

## SURVEY RESEARCH IN THE COURTS: IS IT LEGAL OR SCIENTIFIC?

Carl C. Hoffmann, Hoffmann Research Associates

The earliest indication that I have been able to find that a survey was submitted as evidence in a court case comes in 1880 in a case in Pennsylvania entitled Carroll v. Ertheiler, [(1880, CC Pa) 1 F 688, an action for trademark infringement]. The defendants appeared to have circulated a set of printed affidavits prepared without seeing the witnesses and "sent them [the affidavits] over the country to be signed by those who might be found willing to sign them. The court went on to say that the method of obtaining testimony was not worthy of encouragement."

For over one hundred years there have been matters before the courts for which the opinions of large numbers of people are considered material and relevant. However, the courtroom is at best an inefficient if not entirely biasing and artificial form of gaining that information. It is not possible to call into a courtroom a representative sample of a universe subject to controversy with sufficiently small error tolerances. This effort could entail calling 500 to 2000 people. The estimated numbers required are used only to insure accurate and precise estimates, not to elaborate new facts. Most of their views would be redundant and the judge and/or jury would not have the patience to listen to direct and cross examination of each witness even on the simplest matters. Further, it is particularly silly to think that a courtroom environment would produce unbiased responses to unambiguous questions, uninfluenced by the atmosphere of the proceedings or the questioning, since it is the goal of both the examiner and the cross examiner to obtain the truth through a dialectical process.

At trial the questions and responses are not the measurement instrument as they are in a survey, but the judge and/or jury who sifts and winnows the information presented for the truth become the measurement instruments. Judge and jury adjust what they hear for supplemental facts, the demeanor of the witnesses and lawyers, and the admittedly biased presentation of each side's case (bias here not necessarily meaning false, just not complete). Court proceedings are good models for data collection and analysis, but they are different from what we as surveyors do.

The court proceedings are a place where each side purposely draws a nonrandom group of witnesses through a structured set of responses with the goal of making a point to benefit the advocate presenting the witness. Questions are biased, purposeful, and oftentimes deliberately ambiguous. In most cases, witnesses speak the truth from their view but not the whole truth of the contrasting sides or there would be no trial. While they generally produce seriousness and candor, it is usually a truth which may be only a partial truth and a seriousness which is often so concentrated as to be myopic.

These factors are tested under cross examination which again is designed purposefully to challenge the sincerity, narrative ability, perception and memory of the witness. (In survey

research, these terms translate to reliability and validity). The cross examination process often produces statements which modify claims to a greater or lesser extent or completely contradict direct testimony. It then becomes the function of the trier of fact to pick out the truth (validity) of the statements, the relevance of the statements to the case at hand, their importance to the major issues of the case, their consistency (reliability) both within the testimony of the individual and the consistency with other evidence being presented.

The judge and/or jury then are the measurement instruments as well as the testers of the theoretical model of associations of the various facts. They must adjust for the humanity, imprecision and artificiality of the setting to obtain a view of what they think to be true. It is clear that in antitrust, trademark, unfair competition, securities fraud, naturalization, full disclosure, damage and personal injury, discrimination, labor relations and even libel cases that court proceedings are at best an inefficient method to collect the large amount of data required. Consequently, the courts have turned increasingly to survey research for answers, adopting the survey researcher's method of data collection, measurement, and analysis in place of the courtroom model.

What follows is an effort to differentiate what survey researchers do to arrive at the truth from what the courts do. While there are distinct differences in the demands and methods of the law from that of survey research, the differences are really ones of focus in the methods by which "truth" is obtained. Further there is much that both sides can learn from one another. Finally, the Federal Rules of Evidence may not be adequate as guides to the admissibility of evidence, for on the one hand, they are too strict, and on the other, too lenient.

In any survey or poll, the survey researcher attempts to ensure the truth, sincerity, candor, memory and narrative ability of all those interviewed. But rather than assuming an adversary model as the court does between two or more contentious parties where ambiguity and bias are necessarily a part of the proceedings, the survey researcher attempts to minimize these factors by design. They insure relevance by selecting the appropriate universe; they insure lack of bias, accurate and precise results by proper sampling design; and they insure lack of bias and clarity by selecting the appropriate data collection techniques, questionnaire design or data gathering instruments. They are careful to articulate the design and to test the reliability and validity of their instruments, thus insuring reproducibility. Rather than relying on intuition or impression, they systematically translate the data gathered into a form that can be statistically analyzed and can be constructed into valid and reliable indicators of factors relevant to the case which can then be inter-related to test the crucial questions related to the case.

## Legal Demands for Survey Research

Survey evidence has been admitted to court largely with respect to three hearsay exceptions. The first involves the measurement of a present sense impression [Federal Rules of Evidence Rule 803.1] or presently existing state of mind [Zippo Manufacturing Company v. Rogers Imports, Inc., 216 F. Supp. 670 (1963)]. Second, it is admissible if once having established equivalent circumstantial guarantees of trustworthiness (from a survey researchers point of view, the survey is scientific), the court determines that:

(a) the statement is offered as evidence of a material fact; (b) the statement is more probative on the point for which it can be offered than any other evidence which the proponent can procure through reasonable efforts; and (c) the general purposes of these rules and the interests of justice will best be served by admission of the statement into evidence. However, a statement may not be admitted under this exception unless the proponent of it makes known to the adverse party sufficiently in advance of the trial or hearing to provide the adverse party with a fair opportunity to prepare to meet it, his intention to offer the statement and the particulars of it, including the name and address of the declarant. [Federal Rules of Evidence, Rule 803.24]

Or third, under Rule 703 of the Federal Rules of Evidence:

The facts or data in the particular case upon which an expert bases an opinion or inference may be those perceived by or made known to him at or before the hearing. If of a type reasonably relied upon by experts in the particular field in forming opinions or inferences upon the subject, the facts or data need not be admissible in evidence. [Baumholser v. Amax Coal Company, 630 F.2d. 550 (1980)].

The three exceptions are quite different but the assurance of survey research trustworthiness connected with all three ought to be the same, i.e., the survey should be scientifically sound. A survey to support the first exception, the present sense impression, is fairly easy for it involves little depth and does not seek to establish the truth of the statement. In other words, if a consumer feels that two cigarette lighters are the same, it is the establishment of the confusion which is important, not the fact that the lighters are the same.

The second exception is obviously the broadest and its standards for trustworthiness ought to be rigidly applied, but as we shall see in the second example presented, we are interested in going beyond present sense impressions in many surveys presented in court to assessing the historic behavior and attitudes of individuals in a great deal of depth.

The third exception is to me the most troubling, for there are many studies, publications and "expert reports" which could not meet the standards of trustworthiness set down by the courts, yet are nonetheless relied upon by "experts." It seems that in this instance, rather than accepting the high standard all sur-

vey researchers should adhere to, we are faced with the lowest common denominator rule, where poorly done research, negligent in standards, will be accepted merely because someone with a suitcase more than 100 miles from home says to a judge he always does it that way.

The confusion of the different exceptions can be best illustrated by the comparison of two court cases, both of which reject the studies performed, but one of which accepts the expert's opinion as evidence.

Perhaps the best example of how not to do a survey for court is to be found in the Pittsburgh Press Club case (PPC). The decision states:

The PPC survey clearly cannot meet the standards specified in the Handbook of Protracted Litigation or in Zippo. It was not scientifically designed; indeed, Dr. Kenkel had never before taken a poll. The sample consists of 281 responses, while large in proportion to the universe of 815 challenged banquets, was not designed to be representative. The respondents, who were all Club members and thus interested in the litigation, were told the precise nature of the litigation and the purpose of the survey. They consequently knew which responses would be helpful to PPC, and conversely, which would be harmful. Moreover, it was possible that a recipient of the questionnaire would fail to respond because he knew an honest response would be harmful to the Club's position. Thus the respondents might have contained a higher percentage of those who could answer in a way helpful to the Club.

It therefore appears that PPC's survey suffers from a severe dearth of any circumstantial guarantees of trustworthiness. Both hearsay dangers--faulty memory and insincerity--loom large in the Club's poll. The respondents were not being asked for a present impression; rather they were being asked for details about banquets which had taken place many years before. In fact the covering letter, recognizing that the members' memories might be "dimmed or imperfect," nonetheless asked for a "best recollection." The respondents were all interested in PPC's prevailing in the lawsuit. Yet they were expressly advised about the nature of the litigation and the survey, as well as which answers would benefit the Club.

[Pittsburgh Press Club v. United States, 579 F.2d 751 (1978)].

The one objection I have with the judge's ruling is that the surveyor was in no way attempting a present sense survey, rather he was really attempting admissibility under the second exception, but executed it poorly.

In another case called Baumholser v. Amax Coal Co., the judge criticizes a geologist who knew nothing of survey research, who in conjunction with the lawyers and parties of the suit constructed a study, used people subject to the suit as interviewers and produced study results and expert testimony which were accepted by the District Court. The Appeals Court criticizes the study and agrees that it was not performed independently of the proceedings and seems to say

that there is no showing that the poll was conducted in accordance with generally accepted survey principles, yet allows the expert to express an "opinion" because of Rule 703. The judge elaborates his reasoning on admissibility:

However, the lack of independent grounds for admissibility does not require reversal. Barnes was testifying as an expert and as such was entitled to rely on hearsay evidence to support his opinion, so long as that evidence was of a type reasonably relied upon by other experts in the field. The evidence need not be independently admissible.

Barnes testified that a similar survey was conducted by the Atomic Energy Commission to study the effects of blasts in Colorado. This testimony was uncontradicted and unrebutted. It more than satisfied the threshold inquiry as to whether other experts would rely upon it. [Baumholser v. Amax Coal Co., 630 F.2d 550 (1980)].

This seems to indicate that if experts rely on cow pies, then cow pies can be served up in court and digested. This raises an issue of what is a negligent study.

#### Differences Between Legal Demands and Survey Research Methodology

The procedures involving trustworthiness of all three exceptions in theory are quite similar, provided experts normally rely on trustworthy survey data. How to insure trustworthiness has been recorded in any number of articles<sup>1</sup> but basically include the following areas of examination.

First, the relevancy and adequacy of the universe must be insured. Here it is generally meant to be the population that is subject to the phenomena being measured. For instance, in the case of cigarette lighters, it would be the population of all cigarette smokers. In the case of potential consumer confusion over the design of a bilge pump in the after market, it would be the purchasers or potential purchasers of a bilge pump. Purchasers could conceivably be divided into wholesalers, distributors, and retailers (if they do not offer both but choose one or the other), or the ultimate purchaser depending upon who is most important in determination of the purchase of the product. It could also mean not just the physical population of individuals but the conditions or the mental process under which they decide which bilge pump to purchase. If the end consumer were the relevant population, this might mean actually observing product confusion in the stores which displayed both products. The universe can be variously defined according to the issues of the particular case as demonstrated in the preceding examples.

Next, a survey will be trustworthy if one can justify the relevancy and adequacy of the sample. Here the surveyor needs to demonstrate the way in which he or she assured representation of a population under the conditions and processes relevant to the case at hand. Sampling is a technique by which one strives for efficiency and cost savings, yet also strives to insure a lack of bias and small tolerances of precision for the values being estimated. The more sophisticated

the sampling process, the more difficult it is to make inferences to the basic population because of the complex weighting schemes required to compute estimates and confidence intervals. Similarly, the more complex the sampling design, the more difficult it is to demonstrate lack of bias in multivariate relationships which, as we shall see in the second example, is increasingly the goal in court of survey research. The sample has to be large enough to insure an acceptable level of error, yet small enough not to artificially create distinctions which are of little relevance to the court. Sample sizes can be quite small as demonstrated in the first example depending on the question.

Trustworthiness may also be insured by the qualifications of the person or persons organizing or managing the poll. To some extent this is a silly measure of quality, since the quality of the survey is to be found in the execution not in the executioner. The survey techniques ought to be the central issue and should be examined carefully to insure that the survey was meticulously planned, executed, and analyzed; that the selection of respondents was according to good sampling techniques and not influenced by people requesting the survey; that the poll was conducted independently from the lawyers--ideally that the survey was conducted without knowledge of the legal proceedings (impossible to achieve in some instances); that the questions asked and techniques used were ones which produced serious, candid, unbiased and unambiguous responses; and that the persons interviewed were allowed the freedom to frame answers in their own terms.

In order to insure trustworthiness, the initial research question, the survey research notes, all planning, background and design information, raw questionnaires, coded data and all programs used to tabulate that data ought to be made available to the opposing side. Further, those who planned, supervised and executed the interviews and tabulated the information should be available for deposition or testimony in court.

This process outlined in numerous law review articles and cases is essentially what all survey researchers are taught in school as being the ideal form of survey design. Its goal is to insure some measure of reproducibility. It is clearly different from the process used in court to determine the trustworthiness and veracity of evidence.

The crucial objection raised by lawyers as early as 1956, and I believe a valid one, to the introduction of survey research on complex legal issues is essentially its antiseptic quality, although it is difficult to avoid the antiseptic in trademark cases.

A chief challenge to the survey method in the future lies not in problems of sampling nor in the ambiguities, candor, and bias of communication, but in overcoming this problem of blandness--this tendency to reduce all opinion to a colorless common denominator. [Blum and Kalven, "The Art of Opinion Research", University of Chicago Law Review, 1956, p. 19]

Essentially the objection is that survey methodology has been unable to avoid the democratic problem of counting each view as one.

Blum and Kalven point out that opinion is held on several dimensions, namely the intensity, the knowledgeability and persistence of the opinion. Yet whether we are talking to individuals about Zippo lighters, about their interpretation of moral character, how long they would have worked had they not been fired, whether they were interested, willing and able to take a job, whether they knew the dangers in advance of using a product, or whether in reading a proxy statement they could have been misled by its wording, we as survey researchers tend to view representativeness in democratic terms. It is clear, however, that we are asking these questions of people who often don't care about such issues, often have never thought about them, and if they have thought about them, their opinions may be at best ambivalent. Many of those interviewed may not be competent to express an opinion on the subject matter being questioned.

Examinations of witnesses in court under direct and cross examination allows the judge and/or jury to hear and watch the responses of the witnesses and to develop their own measures of reliability, validity, competence, persistence, intensity, and sincerity. In the best of circumstances, trial lawyers select witnesses who have these qualities. Admittedly they select a biased sample of witnesses and the representativeness of the numbers of people presented by each side in court is not questionable but deliberately non-representative. Yet the witnesses are clearly on display for these crucial measures. In essence lawyers, by seeking the best witnesses, seek articulate people with informed, intensely held opinions. However, the standard governing judgement of opinions in courts is plurality and not representative (democratic) in the sampling sense.

Further, it is definitely true that neither the judge nor the jury, the "triers of fact," are representative, even though jury panels are drawn to be representative of the population surrounding the court. Through challenges and voir dire, the advocates select a set of jurors in whom both sides attempt to decrease the uncertainty of the jury's decisions. They strive for a set of reasonable and prudent people to hear the case, albeit both sides define reasonable and prudent as someone agreeing with them. Not only is there an attempt made to obtain validity, reliability and strength of opinion in witnesses, but also to obtain the same characteristics among jurors.

#### **Example One: The Reasonable and Prudent Man**

In this first example, we present a case study which will examine how a legal concept modifies the survey researcher's sense of representativeness of the sample. The concept is that of the reasonable and prudent man as used in securities fraud. Here one is interested in whether there was an act which could have deceived a reasonable and prudent person. This is not the population of the people exposed to the alleged deception or even one which might be closely analogous. Our task here is to test whether this theoretical person was in fact deceived. I will present a full case study and we shall concentrate on the

translation of this legal concept into an experimental design. The design is not necessarily the ideal, but is one which is able to be executed in the three weeks that this researcher was provided to execute the study.

The issue of what is materially relevant to a prudent and reasonable man occurred in a securities fraud case which involved a proxy contest at a savings and loan in Atlanta, Georgia. The issue revolved around Section 1489 of the Rules and Regulations of the Securities Exchange Commission. In this case, the Board of Directors was evenly split, and as a consequence there was no clearly defined bank representative. Nonetheless, one side appeared to represent itself as the official proxy solicitation of the bank, while the other side fully disclosed the conditions of the bank. The issue of the survey, in this case the experimental design, was the materiality, weight and importance of the representations in the individual proxy solicitations. What significance would a reasonable and prudent person make of the statements in the proxies? How would those statements affect their decision making?

A universe defined as a reasonable and prudent man's interpretation of a proxy solicitation is different from what a social scientist or survey researcher would immediately address when measuring the impact of a proxy solicitation. A surveyor's choice initially would be to ask a representative sample of stockholders in the savings and loan. Yet by having experienced the proxy battle and its effects, this population is tainted and has been subjected to considerable study effects. The next choice would be to go to a savings and loan whose stockholders were similar in most respects to the ones subject to the proxy fight. However, according to the regulations, even this population is not the appropriate one. For the law recognizes that not all stockholders may be reasonable or prudent and so some effort must be made to establish the reasonableness and prudence of the individuals being asked to evaluate the proxies. In fact, stockholders need not be the class of individuals since it is the reasonableness and prudence of the people that is crucial. It is the possibility of deception that is crucial here, not the population deceived.

Secondarily, it would be wise to at least insure that the reasonableness and prudence of the population being interviewed was similar to the population that would have potentially been stockholders and therefore subject to proxy solicitation. Finally, it is important to measure the impact of the deception.

To demonstrate how different the solicitations were, we will quote from both sides. One side, termed the "King" group, indicated their position as follows:

As described below, Messrs. Hanger and King strongly differ with Mr. Edwards, who has thus far been supported by Mr. Cotsakis, concerning the policies and practices to be followed by the Association. Efforts to resolve those differences have failed, and it appears that the present members of the Board are evenly divided. The Association holds revocable

proxies from most of its members. These proxies can only be voted in accordance with the vote of a majority of the entire Board. In view of the division of the Board, it will not be possible to vote these proxies at the Annual Meeting. Accordingly, the Committee has been formed to solicit proxies to be voted in favor of the Committee Nominees. The Committee wishes to elect a Board of Directors which is willing and able to reverse the present policies and practices of the Association and improve its financial conditions and customer service.

The other side, termed the "Edwards" group, presented the following statement and letter:

STATEMENT OF MANAGEMENT WITH RESPECT TO  
SOLICITATION OF PROXIES FOR THE  
1981 ANNUAL MEETING

This statement is presented by George J. Cotsakis, Chairman of the Board of Directors and David C. Edwards, President Peachtree Federal Savings and Loan Association in connection with the accompanying solicitation of proxies for the 1981 Annual Meeting of the Association.

The King group clearly identifies themselves as two of four members of an evenly split Board of Directors and that the issue involved is a non-routine proxy solicitation. The Edwards group makes a deliberate attempt to say that they are management and in no way mentions an evenly divided Board of Directors. The question that remains is how important and material were these claims and did they affect or would they affect the voting patterns of a reasonable and prudent man. Hence a study was designed to answer the following question: would a prudent man have construed that the Edwards solicitation for proxies was the official solicitation of proxies from the bank? A simple experimental design was used to compare readers' interpretations of the "Edwards" solicitation and the "King" solicitation with respect to this issue. Efforts were made to insure that differing interpretations could be derived from the documents themselves and not from the design of the study or the form of communication.

The question now becomes: would the prudent reader looking at the texts of the two solicitations have construed one more than another to be official? Also in what other ways do the documents differ? We should point out here that we do not know whether either group had the ability to represent itself in the official capacity as bank representatives. We merely ask whether a prudent man looking at the two documents would draw different conclusions from them.

#### Execution of the Study

The first stage in determining the concept of the prudent man was here operationalized as those individuals called for jury duty during the week of November 16th in Fulton County. We must take as a given that the Administrator of the Fulton County Courts attempted to draw a set of individuals whom the court believes representative of prudent men.

All of the forty-two individuals available volunteered to participate in the study. These

individuals were randomly assigned to read either the King solicitation or the Edwards solicitation. The groups were formed by having them draw (blindly) either a white, yellow, blue, or red poker chip from a box that contained 50 of each color chip.

Upon drawing a chip, participants were given a packet that was color coded. All packets contained a set of instructions. The packet for the blue and red groups contained the King letter, while the packet for the white and yellow groups contained the Edwards letter. Participants first were asked to read the solicitation and then to answer the questions. The questions were identical for each group. Participants were allowed to refer back to the document if that was required, but they were asked not to answer the questions out of order. As will be demonstrated, the questions not only assessed the difference in the solicitations, but also questions which can be used to examine reasonableness and prudence.

#### Results

The first question that related to whether or not this document appeared to be official was whether the document gave an impression of representing a regular occurrence or a special solicitation. 56% of the group that received the Edwards solicitation felt that it was a special solicitation, while 9% of the group which received the King letter felt that it was a regular occurrence. This is a statistically significant difference between the groups and could not have occurred more than 5 times in 100 by chance.

The next question involving an "impression of authority" is one which refers to the support of the Board of Directors with respect to the nomination of new persons to the Board. Approximately 74% of the individuals receiving the Edwards letter felt that the Board was unanimously in favor or had a majority in favor of the slate of nominees, while 18% of the group that read the King letter answered similarly. The majority of those who received the King letter felt that the board was evenly divided with respect to the nominations. Again, these differences are statistically significant.

Finally, a question was asked to assess the group context surrounding the document. Eighteen out of twenty, or 90%, of the Edwards group felt the group setting around that document was management, or management and the Board of Directors, or the Board of Directors, or was the official solicitation of the bank, while 27% of the King group stated that this was true. This again is a significant difference.

The above differences may have arisen for two reasons. One document may have communicated badly and thus led to a misunderstanding because of the level or quality of communication. We found no evidence of this. We asked several questions about the quality of communication of the two solicitations and we found no difference with respect to the clarity of instructions, the understandability of the vocabulary, the clarity of the issues, the level of business sophistication required to interpret the document, or the logical presentation of ideas in the text.

Further, there are additional indications that information was presented equally well and equal-

ly clearly by both groups. 82% of the group who received the King solicitation and 70% of the group who received the Edwards solicitation understood that there were four seats to be filled at the annual meeting. This is not a significant difference. Finally, 55% of the King group and 53% of the Edwards group correctly answered the question which asked what terms of office the nominees were to fill. From all of this we conclude that both documents were able to communicate factual information equally well.

This information could also be used as a form of cross examination or voir dire to determine reasonableness and prudence. If, for instance, we limited the analysis to only those individuals who correctly answered the factual questions on the number of seats and the terms of office, we could examine the response differences for only those who are reading the material carefully. Further, by limiting the analysis to those people who report the proxies as being understandable and clear and who would also qualify the respondents in terms of their perception, we could establish the self-reported trustworthiness of the respondents and assess the reliability and validity of the individual responses. This allows us to qualify the opinions held of what is admittedly a present sense impression.

Were there any differences with regard to the background characteristics of the two groups which could have explained the different reactions to the solicitations? The answer is no. There are no statistically significant differences in terms of the percentage having savings accounts, the median age of the two groups, the percent black in each group, the average years of education, or the percent of each group that are white-collar.

Finally, we asked a general question about what source of information respondents would be more likely to trust: an official solicitation or one sent from an outside group? Equivalent proportions of each group (71% of the King group and 65% of the Edwards group) responded that they would be more likely to trust as a source of information an official solicitation. This indicates that for at least the 42 people that we questioned there is a tendency to place weight on official information.

#### Conclusions

We conclude that the Edwards letter is different from the King letter and that the former can be construed as an "official" communication according to our definition. Moreover, the perceived differences between the two letters did not arise either from the style of writing or presentation of the two documents, or from the randomized selection procedures which defined the two groups in the experiment. Finally, we have determined that the preference for official documents as sources of information was not affected by the two different types of solicitations (one of which, the "Edwards" letter, was clearly perceived as coming from management or the Board of Directors).

With respect to the adequacy of the study, there are two issues which should be dealt with. The first is the size of the sample. In this case, the documents are so distinct that the

numbers of people required to form a large enough sample to establish that distinction are quite small. The second is that we are dealing here with an experimental design and so we control many of the conditions under which the stimuli were given by the experimental process and we do not need large samples to establish the statistical controls that are substitutes in survey research for experimental controls. Since the purpose is to establish that two "stimuli" were different, this could be done with a small group. Obtaining a more precise estimate of the actual effects on the population would require a larger sample size.

Arguably, we should have several other measures. For instance, perhaps two more groups of people, one that is provided with the King solicitation first, followed by the Edwards solicitation and another that would reverse that order. Arguably, the real population read the solicitations in this order and so we should take into consideration the interaction process on their decisions. But this is, again arguably, icing on the cake, since the law only requires that one or the other or both of the documents was materially misleading and could be misconstrued.

This experiment demonstrates that one can incorporate into polls or experimental designs one of the aspects that is achieved in court, namely the use of supplemental questions to insure that respondents are reasonable and prudent in their assessment.

#### **Example Two: Unbiased Challenges of the Knowledgeability, Intensity and Persistence of Opinions**

One of the most telling criticisms of survey research, especially that which involves public opinion polls and the use of Likert item scales is that it is superficial and game-like and does not test the knowledgeability, intensity and persistence of attitudes as cross examination does. So one does not get the sense of depth of the belief or opinion as a judge would watching a witness cross examined. The intensity, persistence and knowledgeability of an opinion can be tested and indeed even some historical facts can be demonstrated by the use of surveys.

This case involved an Office of Federal Contracts and Compliance Administrative Proceeding of sex segregation, unequal pay, and comparable worth in which women were concentrated in the filling and packaging area of a pharmaceutical company and men were concentrated in the production area. The skill requirements were approximately the same, although the pay was somewhat higher in the production area. The survey was commissioned to see why those patterns existed and whether or not they were a result of discrimination.

The questionnaire observed a maturation of opinion which the questionnaire was designed to inform but not to lead. The questionnaire contained approximately 450 questions, very few of which were scaled items, most of which were yes or no questions which set up complex sets of contingencies and choices. After the choices were made, the respondents were asked to explain those choices in their own words. Coders were

trained at length over a week long period to content analyze the open-ended responses and test it for interrater reliability. A detailed description of the methodology is available from the author.

Early in the questionnaire, individuals in the packaging department were asked whether they were interested in jobs in the production area. Affirmative answers were received to that question from 59% of the men and 30% of the women. This response, if translated into action, would produce a much larger number of women than the 7% female which has historically been in the production division. At the end of the questionnaire, after a series of questions dealing with the requirements of the job and the individuals' treatment by the company, use of time at home, ultimate aspirations and desires, availability for (and interest in) working long hours, and historical behavior with respect to time and pressure demands of the job, a hypothetical offer was made of a position in the production division to men and women in the packaging department. Only 9% of the females who were offered this position accepted it, whereas 50% of the males accepted the hypothetical offer. These acceptance rates parallel the actual ratio of the sexes in the two departments.

The job-interest responses obtained early in the questionnaire reflect initial, "shoot from the hip" impressions, removed from the consequences of the decision with attitudes expressed prior to full evaluation of all aspects of the decision.

Why did 21% of the women decide they would not accept a job in which they had earlier expressed interest? Why did 14.28% of the men who had initially expressed interest in the job reject the hypothetical offer, while another 14.28% who had initially stated they were not interested decide later to accept? The reasons underlying these changes in attitude become clear as the complex factors elucidated by the survey are explored.

#### Perception of the Job

The people in the filling and packaging division may have misperceived the job demands of the production division. People in packaging may view the production job as more difficult than it is, and so the perception of the job, not the reality of the job, may affect their desire to transfer. These differences in perception would have a disparate impact on filling and packaging generally and females especially, because they dominate that division. Alternatively, men and women may view the two jobs differently. Women may view the production job as much more difficult or unpleasant than the filling and packaging job, while men may view the jobs as essentially the same. The differing perceptions would affect the relative acceptance rates of jobs in the production division.

Majorities of employees in both divisions perceive the production division personnel as experiencing an increase in the number of hours worked and in salary. They see a decrease in the amount of time which would be available for oneself and one's family. Pluralities in both di-

visions also expect an increase in the amount of dirt, noise, danger and extremes of temperature which must be tolerated. They perceive the work to require an increase in the weight which must be lifted and increases in the amount of detailed, mechanical, and complex work which must be done. The only real difference of opinion between divisions is that production employees perceive their work to be less boring and tedious than work in filling and packaging, but only 28% of the employees in that division agree. If one compares the actual job descriptions for the two divisions, one finds that the work in the production division requires somewhat more technical skill in operation of computers, somewhat more detailed work with respect to understanding formulae of mixes and batches, more lifting, worse working conditions, and longer hours than work in the filling and packaging division. Work in the production division also involves rotating shifts and weekend work. Filling and packaging division involves more repetitive assembly line work with fixed shifts that do not operate on weekends, and does not require the same qualifications with respect to mechanical ability. The result is a confirmation that the job descriptions are accurately perceived by the employees when they compare their positions to the others.

A composite score was constructed to measure the extent to which an individual's perception of job changes between divisions compared with the group consensus. Scores range from -12, indicating a totally opposite view from consensus, to +12, indicating total agreement with the consensus view. The survey found no significant differences between males and females within either division. Women's differing perceptions of the requirements of the jobs cannot be blamed for the differences in sex representation in the two divisions.

There is some indication that women's view of themselves is not the cause of their lack of representation in the production division. The attitudes of their supervisors could also affect the willingness of women to transfer into traditionally "male" jobs. If these women believe that the supervisors in the production division held sex-stereotyped views of these jobs, would this inhibit their movement into that division? The answer is no. Roughly equivalent proportions of men and women (8% and 15%, respectively) felt that they would not receive the necessary help from supervisors for successfully performing the jobs in the production division.

But what of the two distinct advantages that men are seen to have over women? To some extent, the company can control the amount of heavy lifting required. Perhaps the work could be automated or the weights of boxes reduced, or the organization of work somewhat restructured. However, the company has little control over the ability of men and women to be away from their families. This is largely determined by the division of labor in the home and, more importantly, the flexibility of men and women in dealing with time commitments at home.

## Utilization of Time Inside the Home

One possible explanation for the attitude that men can be away from their homes and families more than women is that women may be more constrained than men by the requirements of managing the household and family. To measure the time commitment outside work for both males and females, interviewees were read a list of 23 common tasks and activities, ranging from recreation to household business activities. For each item on the list they were asked how often they spent time on that activity on a scale ranging from every day to less often than once a month. These items were grouped into two categories based on how closely correlated each estimate was with the others: a housekeeping group, which includes seven items, and a child care group, which includes five items.

On the housekeeping group, females had an average score of 31.31, compared with 22.07 for males. On the child care group, however, males score higher than females: 14.91 for males compared with 12.31 for females. On the combined housekeeping and child care factor, again females score higher than males: 43.63 compared with 37.04. All of these differences are significant. Women have time commitments related to household responsibilities that men do not have. These commitments may limit the time women have available for work outside the home.

It should also be emphasized that the allocation of time now is being described, not its possible allocation were job demands to be changed. Why men have fewer overall time commitments including housework and child care may in part be due to the relative importance of their jobs.

## The Relationship of Work Inside and Outside the Home

Sex plays an important role in the correlations between unavailability for work and the variables measuring time spent on household and child care tasks. Unavailability for work is measured from a question that asked, "Does your home schedule keep you from working some shift assignment?" The variables measuring time spent on housework and child care are the two factors discussed above.

Men and women are constantly in the position of having to make choices about how they use their time. A person can clean up the house or put in overtime. A person can go to work early and stay late to learn a new job or can go home and take care of the children. A person can take a higher paying job that has rotating shifts or avoid the job in order to take care of one's household responsibilities. The survey shows that there is no relationship for males between their unavailability for work and the amount of time spent housekeeping, the amount of time spent with children, or the amount of time spent on both. This clearly indicates that they are extremely flexible. Their time allocation in the home is less a commitment to the home than the way they spend their spare time. They can choose to work any hours and substitute the work outside of the home for work inside the home. When one looks at females, one sees that the correlation between

time spent in the house and unavailability for work is opposite that of men. For men the relationship is the more time spent at housework the more available they are to work shifts. The more housework women do, the more unavailable they are. These associations are significantly different from one another. The correlation for women between child care and unavailability is ten times that for married men. When you combine both activities, the correlation for women is roughly .5 compared with a -.12 for men.

This indicates that women were very conscious of their household responsibilities when answering the question about work availability. Their tasks in the home are commitments to the home. They translated their duties as mothers and housekeepers into limitations on their availability for work, while married men did not. In fact, there is a slight indication that men may flee home responsibilities through work. The data suggest also that men are much more elastic with respect to the demands of the job. For example, if a rotating shift comes along which pays more money, men are much more able to act on that opportunity than married women with responsibilities at home.

These correlations tend to indicate that men have the primary job because they lack commitment at home. In this facility, women remained in the positions that they did not because of discrimination or feelings of inferiority or inadequate knowledge of the requirements or availability of better jobs or because of different interests or stereotyped views of themselves, but because of the responsibilities placed upon them in their private lives.

## **Conclusions**

The criticisms of social science research advanced by Kalven and Blum 30 years ago are even more true today. Much of social science research in the academic literature is extremely bland and antiseptic. In an effort to avoid biased, leading questions, and in an effort to insure scientifically representative samples, survey research has tended toward the superficial and has lost much of the energy, risk taking and innovation which existed in the mid-1950's. The generalized language of survey research, the randomized order of questioning, the stress of the five-point Likert item scales, the desire to eliminate patterned response bias by asking questions in the negative, and the desire to get "representative" samples of populations cheaply and economically all have contributed to this colorlessness. When one compares the writings of Stouffer, Hyman, Lazarsfeld or Rosenberg to Gallop, Harris or Roper or the CBS/New York Times poll, one is shocked not only by the loss of literary style but also by the loss of substance.

What we can learn from the courts and legal proceedings is that it is possible and indeed relevant to view reliability and validity of statements in terms which are no less scientific but a great deal more substantive than those presently used in the social sciences. It is valid, important and not unscientific to isolate populations who are knowledgeable, reasonable and prudent. To exclude populations as relevant may



be undemocratic but it is not unscientific. The question of representativeness of samples for decision making and for legal purposes is much more complex and requires a great deal of art and subtlety to evaluate than methods currently used in opinion research. This style of survey research is the type performed in marketing and political strategy work which is quite effective in building coalitions among factions who are knowledgeable, who have strongly held beliefs, and who are available to be influenced to purchase or vote for a product. But because of the proprietary interests connected with people who do this research, much of the scientific methodology is unavailable.

I think a marriage between survey researchers and the courts can provide the survey researcher with an appreciation of how cross examination can be used to establish the validity and reliability of responses. Present scaling methods are really inadequate for this purpose. They measure consistency and indirectly measure latent traits rather than directly challenging the attitudes, beliefs or behaviors that have been indirectly measured. As Example Two demonstrated, cross examination and elaboration within a survey instrument can radically change the results of the survey. The issue is how to do this scientifically and in a non-directed fashion. This of necessity entails risks. It will always be true that researchers who choose this form of survey technique are much more open to questions and criticisms from their colleagues. But the results of this type of survey are much richer and elaborative, and are worth the potential criticisms which are usually focused on form and not

on substance. The important thing to recognize is that surveys prepared for the courts will themselves be closely scrutinized by opposing experts and responsible counsel. This process is unavailable in the presentation of opinion research done by Harris, Gallop and Roper.

There is a fear among many survey researchers that the courts are a corrupting influence and ask us to sin against the scientific principles which we know. It is true that the courts are not safe places, and it is also true that they are places filled with temptation, but they have built into them a check against the misuse of research which is missing in almost every other arena. In fact, courts are exciting places to develop and exchange ideas and their techniques ought to be adapted to our profession.

#### NOTES

<sup>1</sup>See for example, Blum and Kalven, "The Art of Opinion Research: A Lawyer's Appraisal of an Emerging Science," 24 University of Chicago Law Review, 1 (1956); Hughes and Hoffmann, "Employee Surveys in Employment Discrimination Litigation," 9 Employee Relations Law Journal, 42 (1983); Fred W. Morgan, "The Admissibility of Consumer Surveys as Legal Evidence in Courts," 43 Journal of Marketing, 33 (Fall 1979); Bill Wishard, "Admissibility of Opinion Survey," 18 POF 2d, 305; Hans Zeisel, "The Uniqueness of Survey Evidence," 45 Cornell Law Quarterly, 322; and "Public Opinion Surveys as Evidence: The Pollsters Go to Court," 66 Harvard Law Review, 498 (1953).