

Testing Expert Opinion via Standard Empirical Methods

Michael T. Motley

“No, I did not have sexual relations with that woman, Ms. Lewinsky.” Although this was not the lie that led to Bill Clinton’s impeachment, it was certainly the claim most widely discussed by the public. Even after the facts became known regarding what behaviors Clinton and Lewinsky did and did not engage in, debate continued over the veracity of Clinton’s denial. He claimed that merely having received oral sex did not qualify as “sexual relations.” Some agreed. Some disagreed and argued that he had lied to the public.

Because connotations are subjective, the veracity of Clinton’s statement is a question of definition, not one of fact. Whether Clinton’s definition is relatively unique or relatively common is, however, an empirical question. It would be fairly easy to test. For example, heterosexual men could complete a questionnaire containing a large number of bogus filler questions to disguise the target questions; which on, say, Item 11 asks participants to specify the number of women with whom they have had sexual relations; then asks on, say, Item 25 how many women they have been with where the intimacy went no further than receiving oral sex; then asks on Item 26 whether the women in Item 25 were included in the

Item 11 count. The higher the number of participants who say no, the more common and therefore legitimate the Clinton connotation, and the more who say yes, the more unique is Clinton's connotation. No version of that empirical test was ever conducted, as far as I know, although it could have been informative. But many legal questions like this one and others in this volume can indeed be tested via empirical tests of the type with which social scientists are intimately familiar.

This chapter discusses the potential to apply basic empirical research methods to issues of courtroom litigation. Essentially, the idea is that usually *expert opinion, although expert, is still opinion; these opinions often represent mere hypotheses; and sometimes these hypotheses are empirically testable.*

Whether or not to actually perform an empirical test is at the discretion of the hiring attorney, of course. On the negative side, from the attorney's point of view, an empirical test will add to the expert's fees. More importantly, it may fail to support the expert's position. (If asked the standard finishing question in a deposition, "Do you have any other information that bears on this case," the expert would have to admit an unsuccessful test of his or her opinion, of course.)¹ On the positive side, good empirical evidence can add considerable credibility and support to the expert's opinion.

Independent of these perhaps obvious pros and cons, the attorney's personal attitude toward science versus opinion may come into play, as well. For example, in the first expert-witness case I ever assisted, I formed an independent opinion consistent with the opinion the soliciting attorney was hoping for. Being empirically oriented, I offered to run a lab experiment to test the opinion. I remember quite well the attorney's response to that suggestion: "No science! You are an expert; your *opinion* is sufficient." I responded that I could not feel as confident under oath with hypothesized opinion as I could with empirically tested opinion and would like to run the test for my own conscience, free of charge. He allowed the test, the results provided overwhelming support of the opinion, he changed his mind and used the data in court, gladly paid for the study, and went on to recommend to colleagues the "expert witness who runs experiments." Since then, about half of the attorneys with whom I have worked have wanted to run empirical tests of the opinion/hypothesis when possible; for various reasons about half have not.

This chapter demonstrates, through examples,² the *variety* of communication-related litigation issues to which empirical methods can be applied. The chapter assumes a basic familiarity with quantitative methods, but keeps it simple nevertheless.

**EXAMPLE 1: IS "GIRL" A DEROGATORY TERM
FOR AN ADULT FEMALE?**

Although it never got as far as the courtroom, I remember being fascinated in about 1978 by a female colleague's formal complaint that a male colleague (not me) continually referred to her and other adult females as "girls." She was quite adamant that "girl" carried negative connotations absent in "woman," her preferred label. The male colleague, once confronted, said that he would try to begin using "woman," but that it would be difficult because it would require changing a lifelong habit. He also excused his lexical choice by claiming that the dialect where he grew up considers "woman" to have the more negative connotation, implying a female of considerably advanced age and reduced vitality.

Although the complaint was being settled reasonably amicably, an experiment was commissioned to compare the connotations of "girl" and "woman." Approximately 30 male and 60 female college students completed a set of 7-point semantic differential scales on written descriptions of four adult individuals—two females, Jane and Joan, for the target data, and two males, John and Jack, for bogus filler data to disguise the purpose of the study. Each description was about five sentences long with different details for each person's college major, number of years since graduation, current job, hobbies, and marital/family status. A manipulation check confirmed that the study had been successfully disguised as an effort to compare impressions of various majors, jobs, and hobbies. In fact, however, there were two randomly assigned versions of the questionnaire. On one, the descriptions of Jane and Joan began, "Jane is a woman who . . ." and "Joan is a woman who . . ."; whereas the other began "Jane is a girl who . . ." and "Joan is a girl who . . ." Otherwise the two questionnaires were identical. (Both John and Jack were "a guy" on both versions, with data for these items ignored.) Subjects rated the four persons described on the following semantic differential scales (with poles randomly ordered): fun-dull, young-old, serious-frivolous, attractive-unattractive, competent-incompetent, and interesting-uninteresting. As for the results, virtually none of the girl-woman comparisons showed a significant difference or even came close. This was the case whether male and female subjects were combined or separated, whether Jane and Joan were combined or separated, and whether the individual semantic differential scales were combined or separated. (The only exception was that female subjects ranked Joan as significantly more interesting— $p < .05$ —when described as a girl.)

In short, the study failed to support the complainant's claim that "woman" has a more positive connotation, and also failed to support her

colleague's claim that "girl" has a more positive connotation. Although it is not clear that either party accepted it as such, the study ended up merely demonstrating the old general semantics adage that meanings are "in people," not "in words" (e.g., Korzybski, 1994).³

This girl-woman experiment has been used as the initial example partly because of chronology, partly for simplicity, and partly because of assumed familiarity with the issue by many readers. Although this complaint was handled internally by the university's judicial process, subsequent examples are from actual court cases.

EXAMPLE 2: WAIVER AND RELEASE— LITERAL CONNOTATION OR "COMMON SENSE"?

Suppose you are the administrator of a high school track meet or band contest or dance recital, and one of the out-of-town student participants shows up and submits the required "Participant's Waiver and Release, but the line marked "Parent's Signature" is blank. You witness the participant's coach, band director, dance teacher, or similar adult representative signing and submitting the form. Do you allow the student to participate because a presumably responsible adult has accepted the conditions of the waiver, or do you interpret "Parent's Signature" literally and refuse to allow the student to participate because the parent is not present to sign?

Essentially, this was the decision facing the director of a wrestling tournament in California in 1992. Knowing that the student wrestler's coach had signed the waiver and release form on the "Parent's Signature" line, he allowed the student to compete. Unfortunately, the student suffered a serious neck injury. His mother sued the wrestling association that had sponsored the tournament, claiming that she specifically had not wanted her son to wrestle in this tournament and had intentionally refused to sign the waiver on the assumption that the absence of her signature on the "Parent's Signature" line would preclude his being allowed to participate (*Johnson v. USA Wrestling, et al.*, 1995).

Unlike most court cases where the meaning of a word or phrase becomes an issue (e.g., see Motley, Chapter 11, this volume), the plaintiff's challenge was not that the waiver was so ambiguous as to allow multiple interpretations, or that the intended meaning was not adequately communicated, but rather the opposite. The challenge was that the phrase, "parent's signature" is so clear and unambiguous that the substitution of the coach's signature should not have been allowed. The response of the defendant (the sponsoring wrestling association and others) was not that the denotative meaning of "parent" is unclear, but rather that the

tournament director's decision to allow the coach's signature to suffice was normal, natural, and reasonable.

A simple experiment with 55 student participants was able to shed a bit of light on the question. Essentially, the participants played the role of tournament director. They were given the following instructions: "Imagine that you are the administrator of a wrestling tournament. Before the tournament, you received [the following] relevant information from the national association that will be sponsoring the tournament." The information included two excerpts from literature provided to the actual tournament directors, as well as a copy of the actual "Participant's Waiver and Release." Together, this constituted about 500 words of fine print, among which were two statements that the form "should be" signed by a "parent or guardian" for wrestlers under the age of 18, and a signature line labeled "Signature (parent sign if under 18)." About half the subjects read this original information; the other half read identical information except that "adult representative" was substituted for "parent."

Participants in both groups were asked to imagine a scenario whereby, on the day of the tournament, (a) one or more wrestlers under age 18 show up expecting to wrestle; (b) they do not have a waiver and release form, but the tournament director provides one; and (c) one or more of them are unable to obtain a parent's signature at this time.

Three kinds of questions were asked of the subjects. First, two questions asked whether it would be consistent with the guidelines if they, as tournament director, allowed the wrestlers to compete with no signed form. Subjects in both groups recognized overwhelmingly that the guidelines called for a signed form. Second, two questions asked whether it would be consistent with the guidelines if they allowed the wrestlers to compete with a form signed by their coach. Overwhelmingly, subjects with the "parent" form said no, and subjects with the "representative" form said yes.

Third, subjects were asked to "Suppose that (a) an under-18 wrestler produces a form signed by his coach, and (b) you know that the form is signed by the coach. Describe what you think *you* would *actually* do in this situation," where Likert-scale responses allowed (a) Would *not* let him wrestle; would hardly consider otherwise, (b) Would *not* let him wrestle, but regretfully, (c) Uncertain, (d) *Would* let him wrestle; but reluctantly, and (e) *Would* let him wrestle; would hardly consider otherwise. This time, the differences were not so overwhelming. Almost half the participants in both groups said that they would allow the student to wrestle, with no significant difference between the groups.

In short, this simple experiment suggested that although a form signed by a coach but not a parent precludes a wrestler's participation

according to virtually anyone's *literal* interpretation of the guidelines for this tournament, nevertheless, many ordinary people (including many jurors, presumably) might very well elect to ignore the guidelines in favor of a more "humane" or "commonsense" decision. The case was settled out of court.

EXAMPLE 3: LIABILITY—PACKAGING IMPLICATIONS?

One of the more popular home exercise devices over the years has been a chin-up bar for use in door frames. Early versions of the device had a suction cup at both ends of a telescoping bar that could be twisted out for a snug fit in the door frame. Contemporary versions have safety brackets that are screwed into the door frame to support the bar.

During the transition between the two styles, a young man was doing chin-ups using a bar without brackets. In order to hang from the bar for a full extension of his arms, he had his legs bent back at the knees. The bar slipped and both kneecaps were shattered when they hit the floor. He sued, saying the bar was unsafe (*Cala v. Best Products*, 1985). The defense responded that metal brackets had been enclosed and were supposed to have been used. The plaintiff claimed that he never saw any metal brackets—that he opened the rectangular box, dumped out the bar, and no brackets (or instructions) came out with it, presumably having been lodged within the box if indeed they were packaged in the box.

It happens that the outside of the box made no mention of brackets. Rather, it only stated the name of the product (Adjustable Door Bar Gym), showed a picture of a man using the bar for chin-ups (with no brackets!), and gave four bulleted statements, as follows:

- Adjustable from 21" to 32"
- Portable and convenient for use at home or office
- Capable of holding up to 200 pounds when properly secured to door frame
- Instructions enclosed. Please read carefully before using product

Apparently, either brackets had always been part of the product but had never been mentioned on the package; or, probably more likely, the product had been sold originally without brackets, accidents were reported, and brackets were added but without altering the box to mention them. One issue in the case was whether the box could have been

easily and inexpensively altered (e.g., perhaps by addition of a simple sticker) in such a way that (a) consumers would have been more likely to use the brackets, and/or (b) would have been more likely to read the instructions, and, (c) less importantly, would have been no less likely to purchase the product.

It was the opinion of an expert witness in communication that either of the two primary goals could have been achieved via any of several modest and inexpensive alterations to the box. To test his opinion, a simple experiment was conducted. The study was disguised as a marketing study to determine "how consumers get certain impressions of a product just from examining the package." Subjects were asked to examine one of seven mock-ups of the Door Bar Gym box—the original and one of six hypothesized improvements ($n \sim 20$ per group)—and then answer two primary questions. One was open-ended: "Imagine that you have just purchased this product, have taken it home, and want to use it for the first time. Please describe what you would do in preparing the product for use. Include briefly everything from the time you open the package to the time you first use the product. Please do not refer back to [the mock-up]." Later, the answers to this question would be examined to see if the subject mentioned reading the instructions and/or mentioned installing brackets, with the number of subjects doing so being compared for the seven versions of the box. The other primary question was a 7-point Likert item (*Definitely would* to *Definitely would not*): "What is the likelihood that you would want to purchase the product [represented by the mock-up]?" Answers to this question were compared across the seven versions, as well.

The six suggested improvements involved editing the original package in one of the following ways:

1. Adding a separate bullet item, "safety brackets and mounting screws enclosed,"
2. Editing the original bullet item, "Capable of holding up to 200 pounds when properly secured to door frame," to include "(Safety brackets enclosed.),"
3. Changing "Instructions enclosed" to "Important safety instructions enclosed,"
4. Adding safety brackets to the picture of the man using the bar,
5. A combination of "1" and "3" from this list, and
6. A combination of "1" and "4" from this list.

When participants' responses to the original box were compared with responses to the six variations, results showed that all six revisions significantly ($p < .01$) increased the likelihood that a consumer would read the enclosed directions before using the bar. (Of participants who saw the original box, only about 40% said that they would read the instructions, whereas for all other groups this ranged from 75% to 90%.) As for mentioning brackets, all six alterations were improvements, three at the .01 level of confidence, one at the .05 level, and two not significantly. Essentially, large numbers of subjects described using brackets if their box mentioned brackets verbally, i.e., "1," "2," "5," and "6" above, whereas fewer did so in the other modification groups, that is "3" and "4" above, *with zero doing so in the original-box group*. As for likelihood of purchasing the product, none of the alterations had negative effect, and one had a statistically significant positive effect (specifically, changing "Instructions enclosed" to "Important safety instructions enclosed").

In short, the results showed that the original box was extremely ineffective in making clear the importance of reading the instructions⁴ or using the brackets, and demonstrated that any of several very simple minor alterations could have made a very big difference in this regard (and without decreasing the attractiveness of the product). The attorney for the plaintiff later said that the experiment played a major role in obtaining a very favorable settlement for his client.

EXAMPLE 4: ADVERTISING—MISLEADING OR NOT?

In a class-action suit claiming misleading advertising, plaintiffs alleged that consumers had been misled by the packaging for the "SoundBlaster" sound card with respect to its processing capacities. (A sound card is a device inserted into a computer to allow, among other things, recording and converting sound input—e.g., from a microphone, CD, etc.—into digital audio files that can be edited and then played back through speakers or headphones.) The plaintiffs claimed that the packaging implied an impressive 24-bit recording, editing, and playback sampling size and 96 kHz sampling rate when in fact the sound card was disappointing for its mere 16-bit, 48 kHz sampling. The defense claimed that some of the sound-card applications did indeed live up to the higher standard; conceded that recording, editing, and playback did not; but claimed that the package said so.

An expert witness in communication filed a detailed report agreeing with the plaintiff that the higher sampling capability was implied for all applications, including recording, editing, and playback. Although the report was solid, it was very lengthy, technical, and contained evi-

dence and arguments that were potentially difficult for jurors to follow, as it outlined more than 20 reasons and ways that the package exaggerated the sound card's true capabilities and/or hid accurate statements of its capabilities. Much easier to understand, presumably, would be data showing simply what the typical consumer's expectations would be upon examining the package.

Accordingly, a study was conducted in which 105 student subjects with above-average grade point averages were asked to play the role of consumer on four products—the SoundBlaster and three “filler” products, the latter included to disguise the study as one on packaging style preferences (rather than false advertising), and to deflect attention off the SoundBlaster package as being of special interest.⁵ For each product, subjects were given a brief description of the generic product, followed by a set of criteria they should look for on a hypothetical shopping trip for that product. Subjects then examined the product's box for as long as they wished in order to determine if the product satisfied the assigned criteria. Finally, subjects answered written questions as to whether the assigned criteria were indeed satisfied by the product, and then repeated the process for the next product.

For the SoundBlaster, subjects were introduced to what a sound card is, including a brief explanation of the concept of sampling size being measured in bits and sampling rate in kHz; and were asked to play the role of a shopper concerned especially about recording, editing, and playback capabilities, and seeking the criteria of 24-bit, 96 kHz processing. They then examined all six sides of the Audigy SoundBlaster box for as long as they wanted (typically about 2 to 3.5 minutes). Then they answered a series of questions as to whether the product would satisfy their shopping criteria.

The first question tested whether they remembered the 24-bit, 96 kHz criteria accurately—97.1% did. The next four questions asked whether the product satisfied those criteria—generally, for recording, for editing, and for playback. Multiple-choice responses identified the product to be exactly the same, slightly better, much better, slightly worse, or much worse than the target criteria. Results showed that as for general specifications, 91.4% thought the product was exactly as specified, 3.8% thought it was slightly better, and only 4.8% thought (correctly) that it was worse. For recording in particular, 95.2% thought (incorrectly) the product met or exceeded the criteria. Likewise for editing (91.4%) and playback (96.2%). Results were virtually identical when subjects were limited to computer science and electrical engineering majors.

The implication, of course, is the same as the expert-witness opinion, namely that consumers were indeed misled by the packaging. That is, they would not be likely to notice the accurate but buried statement

about 16-bit, 48 kHz sampling and would instead be influenced by the several overt references to far more desirable 24-bit, 96 kHz processing. But although the opinion and the empirical test matched, the empirical test was judged to be much more easily understood than the expert-witness report's descriptions of more than 20 exaggerations among more than 570 pieces of information on the box. Indeed, the judge in the case said as much.

EXAMPLE 5: LIABILITY—CLEAR WARNING?

Not too long ago, the vehicle used by most parking patrol officers had three wheels and resembled a golf cart with an enclosed cab. One reason for the more contemporary designs—four wheels in most cases—is a 1987 accident in California. While an officer was driving her three-wheeled cart, a car pulled in front of her, and she swerved to avoid collision. Her vehicle was top heavy and the hard turn caused it to tip over. As it was tipping, she fell out and the vehicle landed on top of her and killed her.

Her estate sued (*Fleming v. Cushman*, 1988). Although most of the liability issues concerned the engineering design of the vehicle—whether it could have been easily designed to withstand hard turns, and so forth—the defense introduced a communication issue by noting that there was a warning sticker on the dashboard that instructed operators to not swerve hard, and suggested that had the warning been heeded, the accident would not have happened.

The warning read as follows:

WHILE OPERATING VEHICLE: Remain seated. Use both hands for steering. Keep arms and legs within vehicle body. Avoid sudden starts and stops. Sudden hard turns can cause upset. Regulate speed to meet road and weather conditions. Do not operate near an explosive environment. If a malfunction occurs, cease operation.

The attorney for the plaintiff thought that “Sudden hard turns can cause upset” was not a sufficient warning of the danger of hard turns, and called a communication scholar for an expert-witness opinion. The expert witness agreed and pointed out that there are several ways that the warning seemed inadequate:

1. It begins with childish and obvious warnings that discourage further reading;

2. The warning to keep arms inside is contradicted by the job of marking tires, thus reducing the credibility of the warning;
3. The warning to avoid sudden stops seems unwarranted for this vehicle and thus is potentially confusing;
4. If, as the defense contended, the operator was supposed to heed the posted warning(s), then the instruction to avoid sudden stops, together with the instruction to avoid hard turns leaves no reasonable course of emergency action except to crash (but there was no seat belt) or jump out;
5. Again, "cause upset" is a weak warning for the potential consequences of hard turns (compared with "roll over" or "flip over," for example); and
6. The "sudden hard turns" warning is buried in the middle of relatively unimportant and unnecessary warnings, instead of being highlighted (e.g., by leading the list).

To test this opinion, an experiment was performed to compare the original warning against a simple modification hypothesized to communicate more clearly. Subjects were randomly assigned to one of three groups, one reading the original warning, and the others reading either of two hypothesized improvements.⁶ Otherwise, the groups' tasks were identical.

Subjects were led to believe that several products' packaging and labeling information was being tested and that they had been assigned a three-wheeled utility vehicle like the one used by parking patrol officers. Subjects read a general description of the vehicle, saw a picture of it, read instructions for using it, including the warning decal; and were given a hypothetical scenario in which they are driving the vehicle when a car suddenly pulls in front of them.

The subjects answered an open ended question, "What do you think you should do in this situation," and completed a 7-point Likert scale (*very safe* to *very dangerous*) on each of five potential courses of action, including "slam on the brakes" and "swerve hard to one side."⁷ Responses were compared between the original warning decal and each of the two hypothesized improvements.

The results were as predicted. On the open-ended question, subjects swerved with the original warning, but not with either modification, the latter preferring overwhelmingly to hit the brakes. And on the Likert items, subjects with either modification rated swerving to be significantly more dangerous than subjects with the original warning.

In short, the experiment provided empirical support for the expert's opinion that the original warning was inadequate in communicating the danger of swerving hard, and that this could have been corrected with minor modifications.

FINAL OBSERVATIONS

Other examples are available, but by now the point probably has been made that some (not all) of the opinions one might be asked to render as a consultant or expert witness can be tested and supported empirically, and that this kind of support can be persuasive in the courtroom. Although all of the examples given here have been from my experience in cases dealing with questions of meaning in messages, it is easy to imagine conducting empirical tests of other kinds of expert opinion, including several of those discussed in this volume. For example, to support an argument for a change of venue, mock-juror subjects could be used to test the effect of pretrial news stories to which actual jurors are likely to have been exposed.

Moreover, although all of the examples given here have had an expert witness testing his own opinion, there are other possibilities. An expert witness with insufficient research-methods training could hire a colleague to conduct a test of his or her opinion, for example.

Although the example experiments discussed here have supported their hypotheses with least at a .05 level of confidence, as is the norm in the social sciences, it is worth noting that less-stringent confidence levels may sometimes suffice in court if necessary. Although I have not yet needed to try it, one can imagine that a jury might be much more persuaded by empirical evidence where there is, say, only a 10% chance that the observed effects were due to chance (i.e., a .10 level of confidence) than they would be by an unsupported expert opinion alone.

In any case, this chapter has introduced an idea that is both novel and highly effective, according to all attorneys with whom I have spoken, but is at the same time very simple, at least for most academicians in the social and behavioral sciences. The idea, again, is that expert-witness opinions often represent empirical questions and often these can be tested empirically, via standard social science research methods, as part of the preparation for a trial. Assuming that the design is solid and the opinions are supported, empirical evidence can boost the confidence of the expert witness, can weaken the opposition's challenges to the expert, and can be a very important piece of evidence in persuading jurors.

NOTES

1. Granted, one could argue that a failed test does not support the null and thus provides no evidence for either side, but this would probably do one's case more harm than good in court.

2. Readers may imagine and prefer different experimental designs for one or more of these examples, of course, as is often the case in social science research.

3. It stands to reason, of course, that in matters of labeling people, the preferences of the one being labeled usually should prevail.

4. As discussed in Chapter 11 (this volume), the instructions communicated about the brackets very poorly anyway.

5. The products were GE Soft White Light Bulb, Sony ICD-B7 Digital Voice Recorder, Audigy SoundBlaster sound card, and Sensodyne toothpaste, in that order.

6. Specifically, bulleted and straight versions (with no significant difference found between the two) of the following: Avoid sudden hard turns, as rollover can occur. This vehicle flips more easily than a four-wheel vehicle when cornering. Avoid sudden starts in confined areas, as collision can occur. Regulate speed to meet road and weather conditions. Remain seated. Use both hands for steering. Keep arms and legs within vehicle body. Do not operate near an explosive environment. If a malfunction occurs, cease operation until condition is corrected.

7. Other courses of action included simultaneously hitting the brakes and swerving, letting off the accelerator and colliding with the car, and jumping from the vehicle before it collided with the car.

REFERENCES

- Cala, P. v. Best Products, et al.* (1985). State of California, County of Sacramento, Case No. 321866.
- Korzybski, A. (1994). *Science and sanity: An introduction to non-Aristotelian systems and general semantics* (5th ed.). Fort Worth, TX: Institute of General Semantics.
- Hayakawa, S.I. (1972). *Language in thought and action* (3rd Ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Johnson, J. v. USA Wrestling, et al.* (1995). State of California, County of Alameda. Case No. 713933-0.