Testing the Admissibility of Trademark Surveys After Daubert

Artemio Rivera
Testing the Admissibility of Trademark Surveys after *Daubert*

*Artemio Rivera* [1]

TABLE OF CONTENTS

I. INTRODUCTION .......................................................................................... 682

II. THE ADMISSION OF EXPERT TESTIMONY UNDER
*Daubert* AND *Kumho* AND THE RECENTLY AMENDED
FEDERAL RULE OF EVIDENCE 702 .......................................................... 683

   A. RELIABILITY .......................................................................................... 685

   B. THE RELEVANCE/“FIT” REQUIREMENT ........................................ 669

   C. RULE 403’S UNFAIR PREJUDICE/CONFUSION PRONG .......... 670

III. THE ADMISSIBILITY OF SURVEYS IN TRADEMARK LITIGATION

   A. WHAT IS A SAMPLE SURVEY? .......................................................... 671

   B. SHOULD FLAWS IN THE DESIGN AND APPLICATION OF
THE SURVEY GO TO ADMISSION OR JUST TO THE
WEIGHT OF THE EVIDENCE? ................................................................. 672

IV. MAIN PRINCIPLES OF A PROPERLY DESIGNED NON-PROBABILITY
SURVEY HOW THEIR VIOLATION SHOULD BE ASSESSED IN
DETERMINING THE ADMISSION OF SURVEYS

   A. THE SURVEY MUST BE DESIGNED AND IMPLEMENTED
BY A PERSON PROPERLY QUALIFIED TO TESTIFY AS AN
EXPERT IN SURVEY RESEARCH .......................................................... 674

---

* B.S.I.E., Purdue University; M.B.A., Inter American University; J.D. University of Puerto Rico; LL.M. Georgetown University Law Center. Mr. Rivera is a trial attorney in San Juan, Puerto Rico. He may be contacted at art@sjlaw.com.
B. THE SURVEY MUST IDENTIFY THE PROPER TARGET POPULATION OR UNIVERSE .................................................. 675

C. THE SAMPLE MUST BE REPRESENTATIVE OF THE TARGET POPULATION ........................................................... 678

D. SURVEY QUESTIONS

1. QUESTIONS MUST BE CLEAR AND UNAMBIGUOUS ............... 680
2. FILTER QUESTIONS ................................................................ 681
3. QUESTIONS MUST NOT BE LEADING ..................................... 683
4. QUESTIONS MUST BE RELEVANT TO THE ISSUES ............... 685

E. WHEN ARE CONTROL QUESTIONS NECESSARY? ................. 686

1. THE SURVEY'S METHODOLOGY MUST APPROXIMATE MARKET CONDITIONS .................................................. 688
2. INTERVIEWERS MUST BE PROPERLY TRAINED AND THE SURVEY MUST BE PROPERLY ADMINISTERED .......... 689

V. CONCLUSIONS ................................................................................................................. 691

I. INTRODUCTION

In Daubert v. Merrell Dow Pharmaceuticals, Inc.1, the Supreme Court held that trial courts must only admit expert testimony in evidence when the testimony is proven reliable and relevant to the issues. Prior to Daubert, most courts conditioned the admissibility of expert testimony upon the general acceptance of the expert's method or theory in her field of expertise, pursuant to the case of Frye v. United States.2 Daubert created a doctrinal change in the law of evidence by moving the focus of the admissibility inquiry from the general acceptability test to a test of reliability and relevance. Since experts in trademark litigation normally base their testimony on survey results3, this doctrinal change directly affects the assessment that trial courts must make in deciding whether to

2 293 F. 1014, 1014 (D.C. Cir. 1923).
3 See Trouble v. Wet Seal, Inc., 179 F.Supp.2d 291, 303 (2001) ("In trademark infringement cases, likelihood of confusion is typically established through expert reports based on consumer surveys.")
accept or not testimony on consumer perception. To be admitted, surveys must conform to the reliability and relevancy requirements of *Daubert* and its progeny.

In spite of *Daubert*, the conventional wisdom in trademark litigation remains that the existence of flaws in the design or implementation of a survey, does not raise an admissibility issue, and instead must only be considered by the fact finder in weighing the evidence. This view is at odds with *Daubert*, because it fails to recognize that the existence of flaws in a survey might affect its reliability and/or relevance to the point of making the survey inadmissible.

Various courts that have recently dealt with survey admissibility have applied *Daubert* and have issued a series of rulings rejecting surveys because of lack of relevance and/or reliability. None of them, however, have engaged in an in-depth discussion of *Daubert* or *Kumho* and their effect on the admission of survey evidence. Motivated by the lack of guidance from the courts, this article delves into the issue of survey admissibility, and analyses how the law of evidence and the general principles of survey research must be accounted when applying *Daubert* and its progeny to survey evidence.

II. THE ADMISSION OF EXPERT TESTIMONY UNDER *DAUBERT AND KUMHO*, AND THE RECENTLY AMENDED FEDERAL RULE OF EVIDENCE 702

Rule 702 is at the core of the admissibility inquiry for expert testimony, and requires expert testimony to be not only relevant, but also reliable. Rule 702, at the time of the *Daubert* decision, provided as follows:

"If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise."

---


5 *Daubert*, 509 U.S. at 589.
Prior to *Daubert*, most courts applied Rule 702 by means of the "general acceptance test". Under such test, courts should admit expert testimony if its theoretical basis is considered good science within the relevant community of experts, and frees the judge from having to assess the scientific reliability of the testimony. The advantage of freeing the judge from assessing technical merits, was, however, faced with the disadvantage of occasionally rejecting novel, but reliable testimony, that may have lacked acceptance only because at the time of trial other experts in the field may have not had the chance to evaluate the theory.

In *Daubert*, two minors and their parents sued Merrell Dow alleging that the prenatal consumption by the mother of a prescription drug named Bendectin caused serious birth defects in the children. The defendant moved the trial court for summary judgment arguing that no published article had ever related the consumption of Bendectin to birth defects. The plaintiffs countered with the testimony of eight well-credentialed experts who concluded that Bendectin could cause birth defects. The plaintiffs' experts based their opinions on test tube and live animal studies, pharmacological studies of the chemical structure of Bendectin, and the "reanalysis of previously published epidemiological (human statistical) studies." The District Court granted summary judgment because the plaintiff's expert testimony did not have the general acceptance necessary for admissibility, and therefore was not sufficient to prove causation. Citing *Frye*, the United States Court of Appeals for the Ninth Circuit affirmed the trial court.

The Supreme Court vacated the Ninth Circuit judgment and remanded the case to the District Court, ordering it to test the admissibility of the plaintiffs' evidence based on its reliability and

---

6 *Frye*, 293 F. at 1014 (D.C. Cir. 1923); Paul C. Giannelli, *The Admissibility of Novel Scientific Evidence: Frye v. United States, a Half-Century Later*, 80 Colum. Law Rev. 1197, 1207 (1980) (In a two-page opinion, the *Frye* Court decided that a lie detector method was not admissible in a criminal case because the use of such method was not widely accepted within the relevant scientific community.


9 *Daubert*, 509 U.S. at 582.

10 *Id*.

11 *Id*.

12 *Id* at 583.


relevance, rather than exclusively upon its general acceptance.\textsuperscript{15} The Court specifically held that FRE 702 superceded Frye\textsuperscript{16}, that FRE 702 does not establish a "general acceptance" requirement, and that instead it requires courts to test expert testimony for relevancy and reliability.

This section will discuss the reliability and relevancy thresholds for expert evidence under Rule 702, as construed by Daubert and its progeny. Since Rule 702 is also the threshold for the admission of non-probability surveys, the following discussion is essential to understand what makes a survey admissible in trademark litigation.

A. RELIABILITY

To be reliable, expert testimony must be based on valid principles of the relevant field of knowledge.\textsuperscript{17} Expert testimony may be tested for reliability by subjecting the expert's method to the four-factor test suggested in Daubert, or to other factors devised by the trial court.\textsuperscript{18} Finding the expert's method to be reliable, however, does not end the inquiry. Rule 702 also requires the expert to apply the method to the facts in a reliable fashion, and to base her opinion on sufficient facts or data. Because of these requirements, courts must inquire about the admissibility of expert testimony whenever there appears to be flaws in the application of the method to the facts, or whenever the data used by the expert seems to be insufficient.\textsuperscript{19}

The text of Rule 702, as it became effective on December 1, 2000, is the following:

If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.

\textsuperscript{15} Daubert, 509 U.S. at 597-98.
\textsuperscript{16} Id. at 587.
\textsuperscript{17} See Kunho Tire Co., Ltd. v. Carmichael 526 U.S. 137, 147-48 (1999).
\textsuperscript{18} Id. at 150.
Fed. R. Evid. 702 (emphasis added to amendments).

To pass muster under Rule 702, expert testimony must be based on the knowledge of the relevant discipline.\textsuperscript{20} Knowledge, as used in Rule 702, "applies to any body of known facts or to any body of ideas inferred from such facts or accepted as truths on good grounds."\textsuperscript{21} In other words, knowledge is the sum of the principles and facts generally accepted as well grounded within a given field, and not just "subjective belief or unsupported speculation."\textsuperscript{22}

Knowledge, however, does not come strictly from the classical sciences. There is knowledge associated with every field of endeavor for which expert testimony may be offered, and in occasions, it may be derived strictly from experience or observation.\textsuperscript{23} The fact that the methodology is novel and perhaps not yet accepted by the relevant community of experts, should not bar the testimony from being admitted if the testimony is based on accepted principles of its field, and thus proves to be reliable.\textsuperscript{24}

The knowledge requirement imposes on the trial judge the duty to ensure that the expert testimony is at least minimally reliable. To assist the gatekeeper in testing the reliability of the proffered expert testimony, the Court in \textit{Daubert} suggested the following four factors: (1) whether the theory or technique proffered \textit{can or has been tested}; (2) whether the theory or technique in question has been "subjected to peer review and publication"; (3) whether the technique has a \textit{known potential rate of error} and \textit{standards controlling the technique’s operation}; and (4) whether the theory or technique has been \textit{generally accepted} within the relevant community of experts.\textsuperscript{25}

The \textit{Daubert} factors, however, were suggested, not mandated. Courts should assess what factors are relevant to test a particular theory on a case-by-case basis, and should be "flexible" in their reliability inquiry.\textsuperscript{26} The gate keeping inquiry of reliability should focus on the facts of the case and therefore may include all, some, or perhaps none of

\begin{itemize}
\item \textsuperscript{20} \textit{Daubert}, 509 U.S. at 592.
\item \textsuperscript{21} \textit{Id.}, citing Webster’s Third New International Dictionary 1252 (1986).
\item \textsuperscript{22} \textit{Id.}
\item \textsuperscript{23} Fed. R. Evid. 702; Jason Lee Holly, \textit{Why the Daubert Standard should Apply to both Expert Opinions Based upon “Technical” or “Other Specialized Knowledge” and to Expert Opinions based upon “Science” or a “Scientific Method”}, 30 Cumberland L. Rev. 247, 257 (2000).
\item \textsuperscript{24} See \textit{Daubert}, 509 U.S. at 587.
\item \textsuperscript{25} \textit{Daubert}, 509 U.S. at 593-94.
\item \textsuperscript{26} \textit{Kumho}, 526 U.S. at 150.
\end{itemize}
the listed factors. Since *Daubert* dealt with scientific evidence, these factors may not be suitable for certain specialized, non-scientific knowledge. In such instances, courts must attempt to employ the *Daubert* factors, but if they are not suitable, they may use their discretion to devise a test that suits the particular technique.

When assessing non-scientific testimony, courts must first attempt to apply the *Daubert* factors, but if the court finds that the use of those factors do not fit the particular knowledge or facts, it may modify the test by rejecting some or all of the factors. Indeed, in *Kumho*, the Court did not test the expert testimony, which was non-scientific, against the *Daubert* factors. Instead, it probed it by asking whether the expert in the case would have used "in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field." Some commentators believe that the flexibility afforded by *Kumho* includes the faculty of testing the reliability of non-scientific evidence by posing questions such as "whether in the real world members of the public routinely turn to this profession for services other than testimony, and whether the expert's assumption rests on a large body of experience in situations quite similar to the fact situation before the court."

---

27 Id. at 150.
29 *Kumho*, 526 U.S. at 152 ("a trial court should consider the specific factors identified in Daubert."); Id. citing Black v. Food Lion, Inc., 171 F.3d 308, 311-12 (5th Cir. 1999) ("In the vast majority of cases, the district court first should decide whether the factors mentioned in Daubert are appropriate. Once it considers the Daubert factors, the court then can consider whether other factors, not mentioned in Daubert, are relevant to the case at hand."); Jason Lee Holly, *supra* note 49, at 270 (suggesting that courts should first attempt to apply the Daubert factors because otherwise experts with less than serious theories may attempt to represent them as non-scientific in an attempt to avoid the rigor of the Daubert factors.)
30 Id.
31 Id; see J. Brook Lathram, *The "Same Intellectual Rigor" Test Provides an Effective Method for Determining the Reliability of All Expert Testimony, Without Regard to Whether the Testimony Comprises "Scientific Knowledge" or "Technical or Specialized Knowledge,"* 28 Mem. L. Rev. 1053, 1057 (1988) ("[T]he least we should expect – and indeed demand– of one called upon to offer sworn expert testimony is that she support her position in the same manner that would reasonably be expected of an expert who, outside the courtroom, adheres faithfully to the standards of intellectual rigor demanded by her profession.")
Despite the flexibility afforded to courts to test for reliability, and the fact that Daubert holds that lack of general acceptance is not by itself dispositive of the reliability inquiry, general acceptance probably remains the most relevant factor of the Daubert test. Judge Becker puts it this way in United States v. Downing33; a case heavily relied upon in Daubert 34:

In many cases ... the acceptance factor may well be decisive, or nearly so. Thus, we expect that a technique that satisfies the Frye test usually will be found to be reliable as well. On the other hand, a known technique, which has been able to attract only minimal support within the community, is likely to be found unreliable.

It must be noted that Daubert did not hold that expert testimony couldn't be admitted solely on general acceptance. It simply held that expert testimony must not be rejected solely on lack of general acceptance. 35 Since the reliability inquiry must be flexible, a court could reasonably conclude that in its discretion the general acceptance of a method proves its reliability.36

Finding that the method used by the expert is reliable is an important step towards admission, but it only fulfills one of the three requirements of Rule 702. Rule 702 also requires courts to ensure that the method be applied in a reliable fashion to the facts, and that the facts upon which the testimony is grounded are sufficient in terms of quantity. These requirements clearly forbid the practice of many courts of avoiding their gate keeping duties when flaws are found in a method, by ruling that flaws only go to the weight of the evidence.37 If errors in the method, or in the application of the method, render the analysis unreliable, then the expert testimony becomes itself unreliable, and therefore inadmissible.38

In a bench trial, a court may choose to avoid the admissibility issue altogether and assess the method's flaws when weighing the probative value of the method's results, because there is no risk of

---

33 753 F.2d 1224, 1238 (3rd Cir. 1985).
34 Stephen A. Saltzburg, ET. AL., supra note 13, at 1225.
35 See Daubert, 509 U.S. at 597-98.
36 See Kumho, 526 U.S. at 152.
37 See Stephen A. Saltzburg, ET. AL., supra note 13 ("Concerns about expert testimony cannot be simply referred to the jury as a question of weight.")
38 Id.
confusing a jury and because courts rarely are reversed for admitting evidence. 39 In jury trials, however, the court's gate keeping duties must be exercised rigorously.

B. THE RELEVANCE/ "FIT" REQUIREMENT

The Daubert factor test was devised by the Court to test the reliability of expert testimony pursuant to Rule 702. Rule 702, however, also requires courts to test expert testimony for relevancy. Thus, testimony that may be found to be reliable under Daubert may still be rejected under Rule 702 for lack of relevancy. The issue of relevancy in this context "is whether expert testimony proffered in the case is sufficiently tied to the facts of the case that it will aid the jury in resolving a factual dispute." 40

Specifically, Rule 702 requires expert testimony to "assist the trier of fact to understand the evidence or to determine a fact in issue." This part of Rule 702 has been held to go to relevancy independently of Rule 402, and has been termed by the courts as the "helpfulness" or "fit" requirement. 41 According to the Court, a finding of reliability for one purpose does not make a particular expert testimony reliable for all purposes. 42

The issue of "fit" often arises in cases in which a party proffers evidence of animal testing to prove that humans became ill because they consumed a drug. In such cases, the fact that a court may hold reliable

39 BIC Corporation and Wipe-Out Products, Inc. v. Far Eastern Source Corporation, 2001 U.S. App. LEXIS 22416, 3-5 (holding that in bench trials "errors in methodology ... properly go only to the weight of the evidence..." " ... the admission of evidence in a bench trial is rarely ground for reversal, for the trial judge is presumed to be able to exclude improper inferences from his or her own decisional analysis.");

Schering Corp. v. Pfizer Inc., 189 F.3d 218, 227-28, 230 (2d Cir. 1999)("[r]egardless of the basis cited for admitting these surveys, errors in methodology go properly only to the weight of the evidence."); Southland Sod Farms v. Stover Seed Co., 108 F.3d 1134, 1143 (9th Cir. 1997); Bacardi & Co. Ltd. v. New York Lighter Co., Inc., 2000 WL 298915, 5 (E.D.N.Y. 2000)(reasoning that Daubert imposes gate keeping duty but that flaws in the methodology go to the weight of the evidence and not to admissibility); A&M Records, Inc. v. Napster, Inc., 2000WL 1170106, 3-4 (N.D. Cal. 2000), aff'd, 239 F.3d 1004 (9th Cir. 2001); Lewis Galoob Toys, Inc. v. Nintendo of America, Inc., 780 F.Supp. 1283, 1296 (N.D. Cal. 1991), aff'd, 964 F.2d 965 (9th Cir. 1992). However, according to one court, most of these cases deal with preliminary injunctions or bench trials, in which case, the best policy is to accept the evidence and then give it the appropriate weight. Simon Property Group, 104 F.Supp. 2d at 1039, n.3 (S.D. Ind. 2000) citing Toys R Us, Inc. v. Canarsie Kiddie Shop, Inc., 559 F.Supp. 1189, 1202 (E.D.N.Y. 1983) (suggesting that in bench trials the "safest course for the trial judge is to admit the evidence and to treat the criticisms as going to the weight of the evidence.");

40 Daubert, 509 U.S. at 591.

41 Daubert, 509 U.S. at 591; Downing, 753 F.2d at 1242.

42 Id.
the methodology used by the expert, by and of itself, does not make such evidence admissible. Even if an expert can reliably establish that the use of a drug on animals causes a particular illness, it does not follow that the same drug will have the same effect on humans. Absent evidence that the results of the animal study can be extrapolated to prove the effects of the drug on humans, the animal study and any testimony based thereupon must be excluded due to lack of fit or relevance. As one court puts it, “animal studies may be methodologically acceptable to show that chemical X increases the risk of cancer in animals, but they may not be methodologically acceptable to show that chemical X increases the risk of cancer in humans.”

Thus, in order to be properly admitted in evidence, expert testimony must not only constitute knowledge, but knowledge for purposes of the case. “Rule 702’s ‘helpfulness’ standard requires a valid scientific connection to the pertinent inquiry as a precondition to admissibility.”

C. Rule 403’s Unfair Prejudice/Confusion Prong

Federal Rule of Evidence 403 (hereinafter “Rule 403”) provides that “[a]lthough relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading to the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence.” Rule 403 is the final threshold to surpass of the gate keeping inquiry. Thus, even if the evidence proves to be somewhat reliable and relevant, if its prejudicial effects outweigh its helpfulness, it must be rejected.

Rule 403 gives courts more discretion when assessing expert testimony, than when assessing lay witness testimony, because of the influential power and misleading potential of the former. In the words of the Court, “[e]xpert evidence can be both powerful and quite misleading because of the difficulty in evaluating it. Because of this risk, the judge in weighing possible prejudice against probative force under Rule 403 of the present rules exercise more control over experts

43 In re Paoli R.R. Yard PCB, 35 F.3d at 743.
44 Id.
45 Daubert, 509 U.S. at 591-92.
46 Jason Lee Holly, supra note 18, at 249.
47 See Id.; Starter Corp. v. Converse, Inc., 170 F.3d 286, 297 (2nd Cir. 1999) (holding that despite its trustworthiness, a survey was inadmissible because it was irrelevant and its probative value was outweighed by its “potential to confuse the issues in the case.”)
48 Jason Lee Holly, supra note 18, at 249.
than over lay witnesses." This discretion, however, is not absolute. Courts should assess the misleading potential of expert evidence only when its reliability and relevancy are weak, and reject only when its probative value is outweighed by its prejudicial effect.

Thus, Rule 403 imposes an additional burden to the party presenting expert testimony, but this burden is not onerous as compared to the reliability and relevance thresholds of Rule 702. Only when reliability and relevancy are weak, should Rule 403 factors be evaluated.

III. THE ADMISSIBILITY OF SURVEYS IN TRADEMARK LITIGATION

The use of surveys has become pervasive in trademark litigation. Their constant use by litigants in trademark cases has led some courts to expect their presentation. To be able to enter a survey in evidence, or to have an expert base her testimony on one, the survey method used must be found to be reliable and relevant pursuant to Rule 702, and the Courts' decisions in Daubert and Kumho. This section discusses the nature of surveys and survey evidence, and how the Daubert factors, the Kumho decision, and Rule 702 affect their admissibility.

A. WHAT IS A SAMPLE SURVEY?

Surveys can be designed to count or measure all of the members of a relevant universe, e.g., all of the employees of a company, or all of the students of a school. Such a design, however, could become too burdensome, or perhaps impossible to implement, if the relevant target population is large and/or dispersed over a large territory. When counting or measuring everybody is too burdensome, surveys use a technique called sampling that allows making an inference about the totality of the universe by counting only a portion of the whole. By sampling, the survey selects, preferably at random, the people to be questioned. Surveys that use the sampling method are often called "sample surveys."

49 Daubert, 509 U.S. at 595.
50 Id; See In re: Paoli R.R. Yard PCB, 35 F.3d at 746-47 (suggesting that evidence that proves to be reliable and relevant under Rule 702 should rarely be rejected on Rule 403 grounds).
52 See J. Thomas McCarthy, 5 McCarthy on Trademarks and Unfair Competition §32:195 (4th ed.).
Sample surveys are usually used in trademark litigation to prove the state of mind of the target population (i.e. the individuals whose perceptions about the relevant marks or advertisement are relevant).\textsuperscript{54} People ("respondents") are asked questions which are recorded, organized, and analyzed according to accepted principles of survey research.

Even though some commentators believe that sample surveys constitute only circumstantial evidence of the state of mind of the surveyed population, many courts have admitted them as evidence of \textit{actual} confusion.\textsuperscript{55} Sample surveys can be very persuasive in litigation because they are normally based upon the interviews of hundreds of people and because they are usually presented through the testimony of an expert.\textsuperscript{56}

\textbf{B. SHOULD FLAWS IN THE DESIGN AND APPLICATION OF THE SURVEY GO TO ADMISSIBILITY OR JUST TO THE WEIGHT OF THE EVIDENCE?}

Before a survey is admitted, the trial court must ensure that its method was applied following the principles of survey research. However, despite the \textit{Daubert} mandate that expert testimony must be reliable and relevant under Rule 702 as a condition to admissibility\textsuperscript{57}, many courts have held that the existence of flaws in a survey's method, or in the application of the method to the facts, does not raise an admissibility issue and may only be used to undermine the weight of the survey as evidence.\textsuperscript{58} Such holdings run counter to Rule 702's requirements that (1) the testimony be based upon sufficient facts or data, (2) that the testimony be the product of reliable principles and methods, and (3) that the witness \textit{applies} the principles and methods reliably to the facts of the case.

The failure to test the reliability and relevance of the "method's application" to the facts could result in the admission of survey evidence of no practical reliability and/or relevance. Simply put, if the method is not properly applied, its results cannot be deemed reliable and/or relevant. Therefore, the possibility of presenting the jury with expert testimony of insufficient reliability and/or relevance makes necessary the

\textsuperscript{54} \textit{Ringling Bros.-Barnum & Bailey Combined Shows, Inc. v. Utah Division of Travel Development}, 170 F.3d 449, 465 (4th Cir. 1999).
\textsuperscript{55} J. Thomas McCarthy \textit{supra} note 48, § 32:184.
\textsuperscript{56} See Edward J. Imwinkelried, \textit{supra} note 27, 2286.
\textsuperscript{57} Stephen A. Saltzburg, \textit{et al.}, \textit{supra} note 13, at 1224.
\textsuperscript{58} See \textit{supra} note 33.
engagement of the court in a Rule 702 inquiry, and proves that it is wrong to automatically admit survey evidence under the premise that flaws merely go to the weight.\textsuperscript{59}

The above, however, does not mean that only perfect surveys should be admitted, or that courts should hold hearings to test every proffered survey.\textsuperscript{60} The threshold for admissibility under Rule 702 is not onerous and therefore, surveys that are reliable and relevant enough to be helpful to the fact finder should be admitted, despite their flaws.\textsuperscript{61} Often the court will correctly determine that the flaws are not substantial enough to render the survey unhelpful and will properly admit the survey, allowing the opposing party to cross examine the expert over the issue of reliability and relevance.\textsuperscript{62}

On the other hand, only when the reliability or relevance of a survey has been put into question, should a court engage in Rule 702 inquiries.\textsuperscript{63} There is no reason to expend time and resources in these inquiries if the parties and the court do not suspect of the survey's reliability and/or relevance.\textsuperscript{64} The reliability/relevance inquiry may also be avoided in bench trials because there is no risk of confusing a jury, and because trial judges are presumed to avoid making impermissible

\textsuperscript{59} Spraying Systems Co., 975 F.2d 387, 394 (7th Cir. 1992) (consumer survey of secondary meaning held to be too flawed to create genuine issue of material fact at summary judgment stage); The Learning Network, Inc. v. Discovery Communications, Inc. 153 F.Supp.2d at 789 ("... it is, and has been even before Daubert, necessary for a district court to pass upon the threshold question of the validity of a survey before submitting the results to a jury for their consideration."). Simon Property Group, 104 F.Supp. at 1039 ("The court need not and should not respond reflexively to every criticism by saying it merely "goes to the weight" of the survey rather than to its admissibility. If the flaws in the proposed survey are too great, the court may find that the probative value of the survey is substantially outweighed by the prejudice, waste of time, and confusion it will cause at trial.")

\textsuperscript{60} The court should inquire about a survey's reliability if in its discretion, there are flaws meriting the analysis, or at least one of the parties challenges the survey's reliability. See Kumho, 526 U.S. at 152. ("The trial court must have the same kind of latitude in deciding how to test an expert's reliability, and to decide whether or when special briefing or other proceedings are needed to investigate reliability, as it enjoys when it decides whether or not that expert's relevant testimony is reliable ... Otherwise, the trial judge would lack the discretionary authority needed both to avoid unnecessary "reliability" proceedings in ordinary cases where the reliability of an expert's methods is properly taken for granted, and to require appropriate proceedings in the less usual or more complex cases where cause for questioning the expert's reliability arises.")

\textsuperscript{61} Simon Property Group, 104 F.Supp. at 1039.

\textsuperscript{62} Daubert, 509 U.S. at 596. ("Vigorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence.")

\textsuperscript{63} See supra note 56.

\textsuperscript{64} Id.
inferences from the evidence.\textsuperscript{65} Courts, of course, have discretion under their gate keeping powers to engage in these inquiries.\textsuperscript{66}

In conclusion, adopting the policy that flaws in the survey go to the weight of the evidence rather than to admissibility, oversimplifies the issue, and runs afoul of the gate keeping duties imposed by Rule 702, and the decisions in \textit{Daubert} and \textit{Kumho}. A trial court should engage in a gatekeeping inquiry whenever the issue of admissibility is brought before it in a trial by jury, and the evidence presented by the parties show the existence of flaws.

IV. MAIN PRINCIPLES OF A PROPERLY DESIGNED NON-PROBABILITY SURVEY HOW THEIR VIOLATION SHOULD BE ASSESSED IN DETERMINING THE ADMISSIBILITY OF SURVEYS

In deciding the admissibility of a survey, a court must decide whether the particular survey complies with the requirements of rules 702 and 403. In other words, the court must decide whether: (1) the accepted principles of survey research have been applied in a reliable fashion; (2) the factual base selected is sufficient to base the expert’s conclusions; (3) the method as applied will produce relevant conclusions; and (4) whether the probative value of the survey outweighs its prejudicial effects.

In order to make these decisions, the principles of survey research, relevant to an admissibility inquiry, must be discussed and analyzed in the context of \textit{Daubert} and its progeny, and rules 702 and 403. Following is a discussion of some of these principles, and the effects that the absence or misapplication of these principles may have on the admissibility of non-probability surveys.

A. \textsc{The survey must be designed and implemented by a person properly qualified to testify as an expert in survey research.}

According to Rule 702, only witnesses qualified as experts “by knowledge, skill, experience, training, or education” may render their opinions in court about matters related to their fields of knowledge. A proper background in education, knowledge, or experience is necessary.


not only for qualification purposes under Rule 702, but also is relevant to the issue of reliability under the same rule.\textsuperscript{67}

Therefore, the ideal expert to design and implement a survey for trademark litigation would be someone with a graduate degree in marketing or statistics, who has taken courses in survey research methods, sampling, measurement, interviewing, and statistics.\textsuperscript{68} The expert must also demonstrate to have experience and knowledge of "survey methodology, including sampling, instrument design (questionnaire and interview construction), and statistical analysis."\textsuperscript{69}

B. THE SURVEY MUST IDENTIFY THE PROPER TARGET POPULATION OR UNIVERSE.

The main use of surveys in trademark litigation is to provide evidence of consumer perception with regards to some or all of the issues before the court.\textsuperscript{70} For the survey to be relevant, the proper population must be selected.\textsuperscript{71} The survey must establish whose perception is relevant, i.e., which groups of people, among the public at large, are likely to make decisions based on the marks or advertising in question.\textsuperscript{72} Surveys that fail to include part of the relevant population (under-inclusive) and surveys that include more than the relevant population (over-inclusive) introduce bias and fail to produce the proper factual basis for the results. If these flaws are significant, the survey results may become irrelevant for failing to fit the relevant issues or be supported by the proper factual basis.\textsuperscript{73}

A survey may be defined as "an attempt to derive from a part of the universe, facts which can be reasonably expected to apply to the entire universe."\textsuperscript{74} In determining the relevant universe to be sampled, the main issue is to decide whose perceptions are relevant.\textsuperscript{75} To make that decision, attorneys and their experts must research the available case law to find out what are the demographics and qualities of those who should be surveyed. The relevant universe may be obtained by narrowing the people at large by the type of purchasers involved, geographical area,
buying habits, or by criteria chosen to obtain the impressions of the people whose state of mind are at issue.\textsuperscript{76} Take for example the case of a plaintiff who owns a federally registered trademark for men shoes with national renown, and sues for trademark infringement a competitor who sells only in the state of Indiana. In this case, the relevant universe may be limited to: men consumers (instead of for example, corporate buyers, or women, or children), who are residents of the state of Indiana\textsuperscript{77}, and that have purchased shoes during the last year or plan to purchase shoes during the next six months.

Being careful at selecting the relevant universe and researching the case law, however, is no guarantee that the court will be satisfied with the survey’s universe selection. Convincing the court that the correct and relevant universe has been selected may be a challenging task, so the party proffering the survey must be ready to argue that the survey, despite any flaw in the universe selection, remains relevant enough to be helpful to the jury, and therefore should be admitted. \textit{Gillette Co. v. Norelco Consumer Prods. Co.} is a good example of the problems that may arise when a court finds flaws in the universe selected by a survey.\textsuperscript{78}

In \textit{Gillette}, the court found a survey unreliable for failing to include women as prospective purchasers of shaving razors.\textsuperscript{79} The court determined that market research showed that women were likely to purchase razors as gifts for men.\textsuperscript{80}

The relevancy of the selected universe might be critical to the survey’s admissibility. It has been held that a survey that selects an irrelevant universe becomes itself irrelevant.\textsuperscript{81} \textit{Daubert} and its progeny have made clear that expert testimony must be properly grounded on the relevant facts of the case and that the conclusions reached by the expert must be relevant to the issues before the court. According to \textit{Daubert}, rules 702 and 402 proscribe experts from taking leaps in their conclusions that are not supported by the facts or data upon which their methodology is grounded. An expert, for example, makes a proscribed leap by basing upon animal studies the conclusion that a certain drug causes cancer in humans, when he lacks evidence that those animal

\textsuperscript{76} Id.

\textsuperscript{77} In an infringement suit, people living in the area or areas where both the infringing and the infringed products are sold usually compose the universe.

\textsuperscript{78} 69 F.Supp.2d 246, 262-63 (D. Mass. 1999).

\textsuperscript{79} Id.

\textsuperscript{80} Id.

\textsuperscript{81} \textit{Home Box Office v. Showtime/The Movie Channel}, 665 F.Supp. 1079, 1083 (S.D.N.Y.), aff’d in part & vacated in part, 832 F.2d 1311 (2d Cir, 1987).
studies provide results that can be applied to humans. Likewise, an expert that selects a universe that is irrelevant to the issues would make a proscribed leap by relating the survey's results to the relevant issues.

When the selection of a universe is substantially different from the relevant universe, the survey will become irrelevant to the issues by not representing the state of mind of the relevant universe. A flaw of such magnitude makes the survey inadmissible for being irrelevant under rules 702 and 402.

However, most of the time the selected universe is not just irrelevant, but either over-inclusive or under-inclusive. The problem with an over-inclusive universe is that it fails to reflect the state of mind of the relevant universe by including the impressions of people whose opinions are not relevant to the issue before the court. Conversely, the under-inclusive universe fails to reflect the relevant universe by failing to include the impressions of people whose state of mind are relevant. In Winning Ways, the court held that in a case where the perception of sporting goods consumers was relevant, the survey failed to select the proper universe by surveying only at college bookstores and excluding sporting goods stores. The court also concluded that the use of the under inclusive universe made the survey inadmissible.

Despite their flaws over-inclusive or under-inclusive surveys may remain relevant and reliable if the universe selected covers a substantial portion of the relevant universe. Over-inclusive surveys can be corrected if the non-relevant part of the universe can be deleted from the data, and the remaining part is sufficient to properly represent the relevant universe. Under-inclusive surveys may also be admitted if they prove to cover a substantial portion of the relevant universe, but

---

82 Shari Seidman Diamond, supra note 49 at 240.
84 Id.
87 Id.
88 Shashank Upadhye, supra note 66, at 596.
pose a tougher challenge since it is impossible to precisely find out how the deleted section of the relevant universe may have affected the survey results.\textsuperscript{89}

The case of an \textit{under-inclusive} universe may be argued as a flaw that affects all three prongs of Rule 702. It fails prong (1) because the under inclusive survey is not based upon sufficient data or responses, to support a relevant conclusion; and prong (3) because an under-inclusive universe may be interpreted as a failure to apply in a reliable fashion the sampling method to the facts of the case. Finally, the under inclusive universe may be construed as a relevancy flaw under the helpfulness prong of Rule 702, because of its failure to represent the relevant population.

On the other hand, the case of an over-inclusive target population may be understood as a flaw that goes to prongs (2) and (3) of Rule 702, as well as to its helpfulness prong. The over-inclusive universe affects the survey's reliability and relevancy the same way the under-inclusive universe does, with the mentioned exception that the over-inclusive universe may be easier to fix by excluding the non-relevant responses from the results. It should be kept in mind, however, that a survey must not be excluded solely because of a finding of over-inclusiveness or under-inclusiveness, if the state of mind of the relevant population can be reasonably approximated from the universe selected.

C. THE SAMPLE MUST BE REPRESENTATIVE OF THE TARGET POPULATION.

Sampling is the procedure used by surveys to collect information from the relevant universe or target population. A properly sampled survey will interview people who “... adequately represent the opinions which are relevant to the litigation.”\textsuperscript{90} To accomplish this, the sampling procedure must be reliable and valid (relevant).

Reliability in the context of sampling procedures means “the ability to get the same data values from several measurements made in a similar manner.”\textsuperscript{91} Accordingly, “... if a survey yielded the same distribution of data from one sample of a given size to the next, when they were drawn in the same way, then there would be relatively low 'sampling error' and high reliability over samples.”\textsuperscript{92} On the other hand, the validity or relevance of the sampling procedure is directly proportional to the

\textsuperscript{89} Id.
\textsuperscript{90} Dick's Sporting Goods, Inc., 188 F.3d at 5.
\textsuperscript{92} Id.
degree to which the survey measures what it is supposed to measure.\footnote{Id.} To be relevant, the sampling procedure “must not be affected by extraneous factors that systematically ‘push’ or ‘pull’ the results in one particular direction. To the degree that things, other than those being measured, affect the results by introducing systematic bias, the results are less valid.”\footnote{Id.}

Errors in the selection of the place and time in which the information is collected may lead to bias, reducing thereby the relevance or fit of the survey. For example, sampling only at a mall near an exclusive community may provide results that are biased towards the perceptions of rich people, whereas sampling in malls only during working hours, may cause bias in favor of the state of mind of unemployed people.

Issues regarding the correctness of sampling procedures fall under the reliability provisions of subsections (2) and (3) of Rule 702, and/or under its the relevance or helpfulness prong. The trial court, when assessing the reliability and relevancy of the sampling procedure, should look into (1) the sufficiency of the sample size (“[m]ost litigation surveys employ samples of 200 – 300 consumers – more if there is a control group, fewer if the relevant universe is very rare ...”\footnote{Thomas D. Dupont, \textit{Use of Surveys and Survey Experts in Trademark Litigation}, in \textit{Practical Tips on Trademark Litigation} 187, 195 (A.B.A., Sec. Intell. Prop. L., 2001).}); (2) the nature of the sampling method (e.g. probability vs. non-probability sampling; telephone vs. mall interception surveys\footnote{Most trademark surveys use the non-probability/mall interception sampling method. Even though the non-probability method does not use a mathematical model, the fact that it can be performed at the place where consumers often make their purchasing decision (malls/stores), and that it is generally accepted by marketing professionals and social scientists, have made it the preferred method by trademark litigants.}); and (3) the application of the sampling method to the data collected by the interviewers.

In order to reduce bias, experts must sample at different locations, hours, and days, and must effectively train their interviewers to limit their discretion in the selection of respondents.\footnote{Shari Seidman Diamond, \textit{supra} note 49, at 246.} Interviewers must be instructed to screen each potential respondent to ensure that he or she is qualified to participate in the survey.\footnote{Thomas D. Dupont, \textit{supra} note 91.} The criteria for the selection of respondents must be objective and clearly explained to the interviewers (preferably provided in writing).\footnote{Id.} Screening questions must not be
leading and must be drafted so that they do not influence the respondents’ answers in the survey.\textsuperscript{100} Finally, copies of the completed interviews from each respondent, along with copies of the interviewer’s instructions, and the completed screening questionnaires, should be produced to the other party during the discovery period as part of the Survey Report.\textsuperscript{101}

The abovementioned considerations in sampling are directed towards ensuring that the survey results are as representative as possible of the selected universe. \textit{Representativeness} in this context means reliability and relevancy in the context of the gate keeping function. Thus, courts must ensure that the universe selected by the survey is relevant to the trademark issues and that the sampling process has been executed following reliable techniques.\textsuperscript{102}

\textbf{D. Survey Questions}

Survey questions must possess certain qualities in order to provide reliable and relevant responses. To obtain those responses, respondents are interviewed by the use of written questionnaires, and their answers are logged into interview forms. For instance, in the case of a trademark infringement survey, respondents may be shown the allegedly infringing mark, as it appears on its product, and be asked “what company do you think puts out this product?” With the assistance of a few other follow-up questions, the survey may be able to provide evidence of confusion or lack thereof in the relevant population.

\textit{1. Questions Must Be Clear and Unambiguous.}

Survey questions must be clear and use wording that are easy to understand by the average person in the target population.\textsuperscript{103} “When unclear questions are included in a survey, they may threaten the validity of the survey by systematically distorting responses if respondents are misled in a particular direction, or by inflating random error if respondents guess because they do not understand the question.”\textsuperscript{104} The problem with unclear wording is that it causes people to understand the

\textsuperscript{100} Id.

\textsuperscript{101} See id.


\textsuperscript{104} Shari Seidman Diamond, \textit{supra} note 49, at 248.
questions differently, and therefore people end-up answering non-intended questions.\textsuperscript{105} Since factors such as age, sub cultural group, and region can affect the meaning of a word, researchers should use simple language and avoid wording that can have multiple interpretations.\textsuperscript{106}

To test questions for clarity, researchers often perform what is known as “pretests.”\textsuperscript{107} These pretests must be performed before the main survey is administered by sampling a small group of people (regularly from 25 to 75), and should help the researcher identify sources of probable confusion within the questions before the final questionnaire is drafted.\textsuperscript{108}

Confusing questions can affect the survey’s reliability by motivating respondents to guess, and may affect its relevancy by misleading respondents in a particular direction. Therefore, confusing questions may be construed as flaws of relevancy under the helpfulness prong of Rule 702, or as flaws in the method or in the application of the method to the facts under subparts (2) and (3). Whether the flaw is one of method, or of application of the method to the facts, is more of an issue of interpretation than of substance, since the issue depends on whether the court considers the survey questions to be an inherent part of the method, or simply a part of the process of applying the method to the facts. Anyway, either interpretation seems reasonable and should yield the same result.

2. \textit{Filter Questions}

Filter questions are included in surveys to avoid guessing by the respondents. Guessing, in turn, affects the reliability of the survey by increasing the survey’s random error. Many respondents may have not created an opinion about the issues relevant to the survey because of, for example, lack of experience with the products or with the products’ advertising. In such cases, if respondents are not given the opportunity to answer that they “do not know,” they may end-up guessing in order to provide an answer. Guessing may be avoided in these situations by the use of \textit{full-filter} or \textit{quasi-filter} questions.\textsuperscript{109} With a full-filter question, the respondent is asked first if she knows of the subject matter, and only if she does, she is further questioned about the relevant issues; whereas

\begin{itemize}
  \item \textsuperscript{105} Floyd J. Fowler, Jr., \textit{supra} note 99.
  \item \textsuperscript{106} \textit{Id}.
  \item \textsuperscript{107} Shari Seidman Diamond, \textit{supra} note 49, at 248.
  \item \textsuperscript{108} \textit{Id}; AAPOR Code of Professional Ethics and Practices, \textit{supra} note 99.
  \item \textsuperscript{109} Shari Seidman Diamond, \textit{supra} note 49, at 249.
\end{itemize}
with a quasi-filter question, the respondent is directly made the substantive question, but instead is given “don’t know” or “no opinion” as an alternative answer.

For example, in a case of false advertisement in which a survey is attempting to demonstrate that a television commercial falsely claims that the product’s effect lasts longer than it does, the survey may include a full-filter question to exclude responses from people who have not notice the lasting-effect claim. Such question may ask whether the respondent noticed in the commercial a claim about the duration of the product’s effect. If the respondent’s answer is “yes,” then she may be asked what she understood that duration to be. Conversely, if the respondent answers that she does not recall any claims about duration, by answering “no,” the interview may be terminated without further questions.

On the other hand, with a quasi-filter question the respondent in the abovementioned example may be asked the following:

The commercial you just saw claims that:

g. Product X is effective for about 12 hours; or
h. Product X is effective for about 24 hours; or
i. Don’t know.

The advantage of the full filter is that it reduces the chances of guessing by taking away the substantive questions from those who first admit not to be knowledgeable in the subject.\textsuperscript{110} The disadvantage is that it reduces the number of available substantive responses (“underreporting”).\textsuperscript{111} “There is some evidence that full-filter questions discourage respondents who actually have opinions from offering them by conveying the implicit suggestion that the respondent can avoid difficult follow-up questions by saying that she or he has no opinion.”\textsuperscript{112}

The lack of filter questions, by and of itself, may not constitute grounds for a determination of lack of admissibility, unless it can be proven that their want has provoked a substantial amount of guessing, thereby rendering the survey unreliable.\textsuperscript{113}

\textsuperscript{110} Id.
\textsuperscript{111} Id.
\textsuperscript{112} Id.
\textsuperscript{113} One way to prove guessing is to conduct various surveys which only difference is the inclusion or exclusion of filter questions, and proving that only those surveys that include them provide consistent results.
3. QUESTIONS MUST NOT BE LEADING.

Leading questions have the detrimental effect of skewing the results of a survey towards the suggested responses, thereby making the survey invalid or of little probative value.\textsuperscript{114} Surveys must be designed so that the content, order, and structure of their questions do not suggest the "proper" response.\textsuperscript{115}

A clear example of an improper leading question would be to show the respondent a sample of the product bearing the allegedly infringing trademark, and then asking her whether the plaintiff puts out the product. That question is clearly leading, because it suggests the respondent what the correct answer is. However, there are other means to lead a respondent that are not so obvious, and therefore require more attention. One of them is to use multiple-choice questions instead of open-ended questions.

Multiple-choice questions\textsuperscript{116} provide a choice of answers to choose from, whereas open-ended questions allow the respondent to answer without any help as to what the proper answer should be. Courts often prefer open-ended questions because, other things equal, they are less suggestive. However, that does not mean they are always preferable.\textsuperscript{117} On the negative side, open-ended questions tend to produce answers that are more diverse and therefore require further inquiry to determine their meaning.\textsuperscript{118} The need for further inquiry increases the chances of introducing error and bias to the survey, or simply put, adds the risk of including questions that may be objected, thereby decreasing the survey’s credibility. Furthermore, multiple-choice questions, despite being more leading, are easier to administrate and tabulate.\textsuperscript{119}

The determination of whether to use open-ended or close-ended questions must be based on the nature of the answers the survey is attempting to elicit.\textsuperscript{120} "Open-ended questions are more appropriate when the survey is attempting to gauge what comes first to a respondent’s mind, but close-ended questions are more suitable for

\textsuperscript{114} See Universal City Studios, Inc. v. Nintendo Co., 746 F.2d 112, 118 (2d Cir. 1984); Anti-Monopoly, Inc. v. General Mills Fun Group, Inc. 684 F.2d 1316 (9th Cir. 1982), cert. denied, 459 U.S. 1227 (1983).
\textsuperscript{115} Susan J. Becker, supra note 82, at 479.
\textsuperscript{116} Referred to by some in the field of survey research as "close-ended questions."
\textsuperscript{117} Shari Seidman, supra note 49, at 251.
\textsuperscript{118} Id.
\textsuperscript{119} Pamela L. Alreck, Robert B. Settle, supra note 87, at 120-21.
\textsuperscript{120} Id.
assessing choices between well-identified options, or obtaining ratings on a clear set of alternatives.”121 When using close-ended questions, the choice of alternatives provided to the respondent must attempt to be “exhaustive.”122 In other words, the alternatives should include all reasonably possible positions a respondent may take on the issue.123

Another possible issue with questions is that the order in which they are made, or the order in which the alternatives for a close-ended question are presented, may suggest the response in a particular direction.124 For example, if the questionnaire mentions the name of the relevant trademark and then in a subsequent question asks: “who do you think puts out this product?” the respondent will be motivated to name the trademark she was previously mentioned.125 For that reason, questions must be made from the general to the specific, so that the specific questions do not suggest the answer to the general questions.126

McCarthy provides an example of a question format often used in surveys to prove confusion as to sponsorship or source of a product:

The respondent is shown the accused product or advertisement and is asked: “What company do you think makes this product?” Responses naming the senior user evidence actual confusion as to source. Respondents who did not name the senior user are then asked a second question: “Do you think this product was approved, licensed or sponsored by another company or not?” If the answer is yes, respondent is asked: “What company do you think this product is approved, licensed or sponsored by?” Responses naming the senior user evidence actual confusion as to sponsorship, affiliation, or connection, which is actionable. Both questions should be followed up by the important question: “Why do you say that?” Often, an examination of the respondent’s verbatim responses to the “why” question are the most illuminating and probative part of a survey, for they provide a window into consumer thought processes in a way that mere statistical data cannot.

It is not leading to posit the possibility to respondents that there may be some form of affiliation or licensing behind a junior user’s operation.127

121 Id.
122 Id. at 123.
123 Id.
124 Id. at 117.
125 J. Thomas McCarthy, supra, note 48, § 32:172.
126 Shari Seidman Diamond, supra note 49, at 254.
127 J. Thomas McCarthy, supra note 48, § 32:175(citations omitted).
Likewise, the order of presentation of the alternatives to a close-ended question can affect the response.128 "In mail surveys, respondents are more likely to select the first choice offered, (primacy effect), while in telephone surveys, respondents are more likely to select the last choice offered (recency effect)."129 To control for order effects, the order in which questions are made and the order in which alternatives to close-ended question are presented, must be rotated among the questionnaires handed to the interviewers.130 This can be accomplished by providing the interviewers with, for example, three or four different questionnaires in equal amounts, which only difference is the order of the questions and/or the order of the alternatives to those questions.131

The use of leading questions can affect the relevancy or validity of the survey by skewing the results in the desired direction. In those circumstances, the court must decide under the helpfulness prong of Rule 702 whether the survey, despite the leading question or questions, remains helpful to the fact finder to understand the evidence or to determine a fact in issue.

4. Questions Must Be Relevant to the Issues.

The requirements of rules 702, 402 and 403 that expert testimony be relevant and not unduly prejudicial mandate that survey questions also be relevant.132 Relevancy in the context of survey questions will mainly depend on whether the questions made to the respondents address the issues or the facts relevant to the case.133 Questions that merely attempt to associate a word or phrase to a product or that merely attempt to prove that one product copies another, are not probative of likelihood of confusion or trademark infringement.134 The reasoning behind this doctrine is simple: if the questions asked in a survey are irrelevant, so is the survey.

128 Pamela L. Alreck, Robert B. Settle, supra note 87, at 117-18.
129 Shari Seidman Diamond, supra note 49, at 254.
130 Id.
131 See Id.
132 Starter, 170 F.3d at 296-97; Simon Property, 104 F.Supp.2d at 1041.
133 Id.; Mead Johnson & Co. v. Abbott Laboratories, 201 F.3d 883, 885-86 (7th Cir. 2000); K'Arson Corp., 166 F.3d at 6; Home Box Office v. Showtime/The Movie Channel, Inc., 665 F.Supp. 1079 (S.D.N.Y. 1987) vacated in part by 832 F.2d 1311 (2nd Cir. 1987);
For example, in *Starter*, the court dealt with the issue of whether the marks the defendant planned to use on an athletic shoe infringed the marks of the plaintiff. Starter Corp. offered a survey in evidence to prove that there was no likelihood of confusion between the marks. The respondents were shown the athletic shoes in dispute among various others, and then were asked to identify which shoe they preferred, and to give the names of the shoes they saw. The court rejected the survey holding that the respondents' preference for shoes, or their memory about the marks they saw, is *not relevant* to the issue of likelihood of confusion.

*Daubert* held that the issues of helpfulness or relevancy of the expert testimony must be decided by the courts under Rule 702. If the court finds that the questions are irrelevant because they are not directed at obtaining answers that may help the fact finder to understand the evidence or determine a fact in issue, then the survey is irrelevant to the case and should be rejected. If the questions are, however, somewhat relevant, then the court should assess pursuant to Rule 403, whether any possible prejudice or confusion that the admission of the survey might cause, outweighs its helpfulness.

**E. WHEN ARE CONTROL QUESTIONS NECESSARY?**

Control questions help define the root of the perceptions reported in a survey. Surveys designed to prove trademark infringement or deceptive advertising usually attempt to prove how a trademark or an advertisement influences the public's perception of a good or service. Such perception may be established by the survey, but the question may remain about the source of the perception. In some instances, without control questions the survey may not be able to properly report whether the source of the confusion comes from a mental association between the marks, or if it is just the result of guessing, natural confusion (confusion

---

135 170 F.3d at 289-91.
136 Because the Starter shoe had not been manufactured at the time of trial, respondents were shown a prototype of it.
137 Id. at 296-97.
138 Id. at 297.
139 509 U.S. at 592 ("Rule 702 further requires that the evidence or testimony 'assist the trier of fact to understand the evidence or to determine a fact in issue.' This condition goes primarily to relevance.")
140 Another aspect of relevancy -is whether expert testimony proffered in the case is sufficiently tied to the facts of the case that it will aid the jury in resolving a factual dispute.") (citations omitted).
140 *Ringling Bros.*, 170 F.3d at 465.
that is deemed to be inevitable and a human condition¹⁴¹), and/or a badly administered survey. The cumulative effect of these factors is known as “general background noise.”¹⁴²

A good example of just how important control questions can be is given in Reed-Union Corp. v. Turtle Wax, Inc. In Reed-Union, the manufacturer of NU FINISH, a car polish product, sued a competitor who sold a similar product under the trademark FINISH 2001. The manufacturer of NU FINISH alleged that consumers were confused about the two marks and were purchasing FINISH 2001, thinking they were getting NU FINISH.¹⁴³ Both parties proffered surveys claiming that approximately 25% of the respondents were confused about the source of the products.¹⁴⁴ The defendant, in addition, included in its survey various control questions about the possibility of confusion between the plaintiff’s mark, NU FINISH, and PRISM (the trademark of another car polish). As a result, approximately 20% of the respondents confused the plaintiff’s mark with the PRISM mark. The court concluded that the confusion reported by 25% of the respondents was therefore mostly caused (20%) by natural confusion, or factors (“noise”) other than the alleged association between NU FINISH and FINISH 2001.

Ascertaining the levels of noise, however, is not the only use of control questions. Control questions have been held necessary in some cases.¹⁴⁵ In Prostyle, the National Football League and the Green Bay Packers football team sued Prostyle for selling unauthorized t-shirts that allegedly resembled the Green Bay Packers Uniform’s shirt.¹⁴⁶ The plaintiffs sued for trademark infringement and dilution, among other claims.¹⁴⁷

Citing Kumho as imposing a “gate-keeping obligation,” the court in ProStyle ruled a survey inadmissible for asking only the question: “What if anything do you think when you see this shirt?” The court held that an answer to that question, without further probing, was meaningless, and that the lack of use of a control question or a control t-shirt, made

¹⁴¹ Reed-Union, 77 F.3d 909 (7th Cir. 1996).
¹⁴³ Reed-Union, 77 F.3d at 912.
¹⁴⁴ Id.
¹⁴⁶ Id. at 666.
¹⁴⁷ Id.
the survey unreliable. More specifically, the court held that the plaintiffs should have shown the respondents a non-infringing control shirt to test whether the respondents connected the allegedly infringing shirt and the Packers because of the shirt's green and white colors, and the fact that the Packers had won a recent championship, or if the connection was truly based on the allegedly infringing design. The Prostyle court held that it was guided in performing the reliability test by “some or all” of the factors suggested in Daubert. It, however, failed to discuss what factors in particular were applicable to the case or how they should be applied.

The lack of necessary controls in a survey may be construed as a failure to use “reliable principles and methods” pursuant to subpart (2) of Rule 702, since a survey that lacks the necessary controls could be accused of lacking a reliable methodology. On the other hand, it can be argued that it constitutes a failure to “appl[y] the principles and methods reliably to the facts of the case”, pursuant to subpart (3) of Rule 702, if one understands the failure of including controls as a flaw in the application of the methodology to the facts, rather than a failure in the methodology itself.

1. The survey's methodology must approximate market conditions.

The context in which the marks are presented to the respondents must approximate the conditions in which consumers are exposed to the marks in the market place. Differences in market conditions can affect consumer perception and therefore should be taken into account in the survey design. For example, if a product is sold in the market bearing its trademark, respondents must be shown a sample that bears the trademark in the same manner. On the other hand, products that usually are not encountered in the market place side-by-side should not be presented that way to the respondents. "At bottom ... a survey to

148 Id. at 668-70.

149 The control should have been a non-infringing green and white football shirt. That way, if the respondents still connected in their minds the non-infringing shirt with the Green Bay Packers, it becomes apparent that the connection is not being caused by the design, but by the consumers' association of the Packers with things that are green and white after the recent championship.

150 Id.


152 W.E. Bassett Co. v. H.C. Cook Co., 164 F.Supp. 278, 285 (D.Conn 1958)(survey results discounted where trademark was not visible on knife shown to respondents, as it was in the market place)

test likelihood of confusion must attempt to replicate the thought processes of consumers encountering the disputed mark or marks as they would in the marketplace.\textsuperscript{154} Because the consumers’ thought process must be as closely replicated as possible, surveys must be administered to respondents when they are in a “buying mood,” which is the reason why surveys are usually conducted in stores or shopping centers.\textsuperscript{155}

A recent example of the importance of approximating market conditions is given in \textit{Simon Property}, in which a survey showed respondents images of the homepages of two competing websites (bearing the marks at issue) almost immediately, one after the other. The court found that on the Internet, homepages are not normally found side-by-side, or immediately one after the other, but by search engines that display their web addresses among many others. The court rejected the survey because in its opinion it failed to reproduce the home pages in the context in which they would normally be found.\textsuperscript{156}

In holding that the survey be rejected, the court stated that “[i]f the flaws in the proposed survey are too great, the court may find that the probative value of the survey is substantially outweighed by the prejudice, waste of time, and confusion it will cause at trial.”\textsuperscript{157} The court in addition explained that it has “a responsibility to the jurors not to waste their time or to make their task unduly difficult by admitting evidence that is likely to be complex and time consuming...”\textsuperscript{158}

\textbf{2. INTERVIEWERS MUST BE PROPERLY TRAINED AND THE SURVEY MUST BE PROPERLY ADMINISTERED.}

The expert conducting the survey must ensure that the interviewers are properly trained in the correct methods of data collection, and must properly supervise the whole survey process.\textsuperscript{159} The main goal of the training stage must be to teach the prospective interviewers to be consistent in the application of the selected data collection procedure.\textsuperscript{160} Standardization of the process should be the goal.\textsuperscript{161}

\textsuperscript{154} \textit{Simon Property}, 104 F.Supp. at 1038 citing J. Thomas McCarthy, \textit{supra} note 48, §32:163 (“the closer the survey method mirrors the situation in which the ordinary person would encounter the trademark, the greater the evidentiary weight of the survey results”).

\textsuperscript{155} J. Thomas McCarthy, \textit{supra} note 48, §32:163.

\textsuperscript{156} \textit{Simon Property}, 104 F.Supp. at 1043.

\textsuperscript{157} \textit{Id.} at 1039.

\textsuperscript{158} \textit{Id.}

\textsuperscript{159} AAPOR Code of Professional Ethics and Practices, \textit{supra} note 99.

\textsuperscript{160} Floyd J. Fowler, \textit{supra} note 99, at 109.

\textsuperscript{161} \textit{Id.}
In order to attain standardization, all interviewers must be trained to perform their job the same way. This requires interviewers to ask the same questions, in the same manner (by reading them), and in the same order.\textsuperscript{162} In addition, interviewers should be provided with documents containing the questions to be made (questionnaires), and should not be told who the sponsor of the survey is.\textsuperscript{163} These measures are directed to avoid the interviewers' styles or prejudices affect the survey results.

For example, if an interviewer is not instructed to read each question the same way to every respondent, she may assume, after having conducted various interviews, that she has memorized the questions, and may begin asking them with certain variations.\textsuperscript{164} The interviewer may also commit the error of modifying the wording of questions depending on the level of education or intelligence she may perceive the respondent to have.\textsuperscript{165} All these modifications may cause the questions to be different, and might have the detrimental effect of eliciting answers different from the ones intended.\textsuperscript{166}

The careless administration of a survey has been held to constitute grounds for inadmissibility.\textsuperscript{167} In \textit{Toys “R” Us}, the court rejected a survey for, among other things, the following reasons: (1) the expert conducting the survey did not know what the interviewers exactly did in conducting the survey, and he could not assure that the interviewers actually followed the instructions they were given; (2) an interviewer who was heard whispering that the survey was sponsored by Toys “R” Us, was not terminated; (3) there was evidence that people who listened to the interviews of others were subsequently interviewed; and (4) no steps were taken to ensure that the people interviewed met the screening criteria.\textsuperscript{168}

Poor training and careless administration of the survey are clearly flaws that go to the expert's duty to apply the method reliably to the facts, pursuant to subpart (3) of Rule 702. The court, as gatekeeper, must ponder whether such flaws are significant enough to render the survey unreliable and thereby inadmissible, or if it remains reliable and relevant enough to be admitted in evidence.

\textsuperscript{162} Id.
\textsuperscript{163} Pamela L. Alreck, Robert B. Settle, \textit{supra} note 87, at 237; Shari Seidman, \textit{supra} note 49, at 232.
\textsuperscript{164} Id.
\textsuperscript{165} Id.
\textsuperscript{166} Id.
\textsuperscript{168} Id. at 1203-04.
V. CONCLUSIONS

Three important conclusions about the admissibility of non-probability surveys can be drawn from Rule 702, as construed by Daubert and its progeny: (1) courts must test surveys for reliability and relevance as a condition to admission; (2) trial courts have discretion in devising the test and in making the ultimate determination of reliability and relevance; and (3) it runs afoul of the mandates of Rule 702, to hold that the existence of flaws in a survey’s methodology, or in the application of the methodology to the facts, goes only to the weight of the evidence, and does not create an admissibility issue.

Rule 702 imposes on trial courts a clear and unambiguous duty to keep unreliable and irrelevant surveys away from the fact finder. Surveys which methodologies or administration are flawed to the point where they cannot reasonably represent the perception of the target population, should be rejected.

In order to be admissible, a survey must apply the principles of survey research to the target population in a reliable manner, and base its results upon sufficient interviews and responses. These requirements make clear that the existence of flaws in a survey is not simply a matter of weight to be resolved by the fact finder, but an issue of admissibility that must be determined by the courts as part of their gate keeping duties.