

The Trademark Reporter®



The Law Journal of the International Trademark Association

Adios! to the Irreparable Harm Presumption in Trademark Law

Anne Gilson LaLonde and Jerome Gilson

Geographical Indications: The United States' Perspective

D. Peter Harvey

Foundational Objectives of Laws Regarding Trademarks and Unfair Competition

Tony Bortolin

What Insurance Coverage Remains for Trademark Infringement Claims After Years of Policy Language Changes?

David A. Gauntlett

Internet Surveys in Intellectual Property Litigation: *Doveryai, No Proveryai*

Himanshu Mishra and Ruth M. Corbin

Book Review: Constructing Intellectual Property, by Alexandra George

Alfred C. Frawley

Book Review: European Trade Mark Law: A Commentary, by Annette Kur and Martin Senftleben, with a contribution by Verena von Bomhard

Edward Vassallo

Book Review: International Trademark Classification: A Guide to the Nice Agreement, by Jessie N. Roberts

Jane F. Collen

INTERNATIONAL TRADEMARK ASSOCIATION

Powerful Network Powerful Brands

655 Third Avenue, New York, NY 10017-5646

Telephone: +1 (212) 642-1733

email: wknox@inta.org

Facsimile: +1 (212) 768-7796

OFFICERS OF THE ASSOCIATION

JOSEPH FERRETTI..... *President*
TISH L. BERARD..... *President Elect*
DAVID LOSSIGNOL..... *Vice President*
AYALA DEUTSCH..... *Vice President*
TIKI DARE..... *Treasurer*
ZEEGER VINK..... *Secretary*
MAURY M. TEPPER, III..... *Counsel*
ETIENNE SANZ DE ACEDO..... *Chief Executive Officer*

The Trademark Reporter Committee

EDITORIAL BOARD

EDITOR-IN-CHIEF, CHAIR

KATHLEEN E. MCCARTHY

STAFF EDITOR-IN-CHIEF

WILLARD KNOX

Senior Editors

JESSICA ELLIOTT CARDON

GLENN MITCHELL

ELISABETH KASZNAZ FEKETE

RAFFI V. ZEROUNIAN

FABRIZIO MIAZZETTO

PAMELA CHESTEK

CHIKAKO MORI

Staff Editor

BEVERLY HARRIS

Staff Editor

JOEL L. BROMBERG

Editors

TSAN ABRAHAMSON
MARIA BARATTA
MARTIN J. BERAN
DANIEL R. BERESKIN
STEFANIA BERGIA
LANNING BRYER
SHELDON BURSHTAIN
ROBERT CAMERON
JANE F. COLLEN
THEODORE H. DAVIS JR.
ANNE DESMOUSSEAU
MEGHAN DILLON
THOMAS F. DUNN
SCOT DUVAL
CLAUS M. ECKHARTT
SHEJA EHTESHAM
KAREN L. ELBURG
MATTHEW EZELL
NEMESIO FERNANDEZ-PACHECO
SALVADOR FERRANDIS
ALFRED FRAWLEY
ALEX GARENS
ALEXANDRA GEORGE
DANIEL GLAZER
ANDREW J. GRAY IV
LESLEY MCCALL GROSSBERG

ANN LAMPORT HAMMITTE
GUY HEATH
ANNE HIARING HOCKING
JANET L. HOFFMAN
GANG HU
DOMINIC HUI
AHMAD HUSSEIN
BRUCE ISAACSON
AGLIKA IVANOVA
E. DEBORAH JAY
FENGTAO JIANG
HE JING
MARIA JOSE JIRON BENDANA
SIEGRUN D. KANE
SUSAN J. KERI
MIKE KEYES
ROLAND KUNZE
JOI MICHELLE LAKES
SCOTT LEBSON
NELS LIPPERT
MARCUS LUEPKE
J. THOMAS MCCARTHY
NANCY A. MILLER
GEORGE W. MOXON
JOHN M. MURPHY
PAUL MUSSELL
SADAF NAKHAEI

SAURABH NANDREKAR
AMANDA NYE
JENIFER DEWOLF PAINE
JEREMY B. PENNANT
NEAL PLATT
MICHEL RIJSDIJK
RACHEL RUDENSKY
JEREMY SCHACHTER
MATTHEW R. SCHANTZ
MARTIN SCHWIMMER
JENNIFER SICKLER
AARON SILVERSTEIN
GIULIO ENRICO SIRONI
WENDI E. SLOANE
JERRE B. SWANN, JR.
SCOTT THOMPSON
CHINASA UWANNA
ANJALI VALSANGKAR
EDWARD E. VASSALLO
MARTIN VIEFHUES
CHARLES WEBSTER
JORDAN WEINSTEIN
JOHN L. WELCH
JOSEPH WELCH
BRYAN K. WHEELLOCK
JOSEPH YANG

Advisory Board

MILES J. ALEXANDER
WILLIAM M. BORCHARD
CLIFFORD W. BROWNING
LANNING G. BRYER
SANDRA EDELMAN
ANTHONY L. FLETCHER
ARTHUR J. GREENBAUM

ROBERT M. KUNSTADT
THEODORE C. MAX
JONATHAN MOSKIN
VINCENT N. PALLADINO
JOHN B. PEGRAM
ALLAN S. PILSON

ROBERT L. RASKOPF
PASQUALE A. RAZZANO
SUSAN REISS
PIER LUIGI RONCAGLIA
HOWARD J. SHIRE
JERRE B. SWANN, SR.
STEVEN M. WEINBERG

The views expressed in *The Trademark Reporter* are those of the individual authors and do not necessarily reflect those of INTA.

The Trademark Reporter (ISSN 0041-056X) is published electronically six times a year by the International Trademark Association, 655 Third Avenue, New York, NY 10017-5646 USA. INTA, the INTA logo, INTERNATIONAL TRADEMARK ASSOCIATION, POWERFUL NETWORK POWERFUL BRANDS, THE TRADEMARK REPORTER, and inta.org are trademarks, service marks, and/or registered trademarks of the International Trademark Association in the United States and certain other jurisdictions.

**INTERNET SURVEYS IN
INTELLECTUAL PROPERTY LITIGATION:
*DOVERYAI, NO PROVERYAI*¹**

By Himanshu Mishra and Ruth M. Corbin** ****

CONTENTS

I.	Introduction	1099
II.	The Need for Internet Survey Standards in the Courts..	1100
III.	Relevance, Reliability, and Validity.....	1101
	A. Relevance	1102
	B. Reliability.....	1102
	C. Validity.....	1103
IV.	Panel Selection and Reliability—Issues and Remedies...	1105
	A. Problems with Internet Panels.....	1107
	1. Identity	1107
	2. Representativeness.....	1107
	3. Motives of Volunteers.....	1109
	B. Risk Management of Internet Panels.....	1110
V.	Participant Response and Survey Validity	1113
	A. Duplicates	1113

1. A Russian proverb that means *Trust but Verify*.

* Dr. Himanshu Mishra is David Eccles Professor of Marketing at the University of Utah. Dr. Mishra's research focuses on consumer decision-making, behavioral decision theory, and consumer psychology. He has served as an expert witness in Lanham Act trademark litigation involving secondary meaning, likelihood of confusion, and deceptive advertising surveys. He may be contacted at himanshu.mishra@utah.edu.

** Dr. Ruth M. Corbin is Chair of the forensic research firm, Corbin Partners Inc., Mediator at Corbin Estates Law, and Adjunct Professor at Osgoode Hall Law School, Toronto, Canada. She is the author of the Canadian Trademark Law Benchbook chapter on social scientific evidence, and chair of the Dispute Resolution Committee of the Market Research and Intelligence Association. She may be contacted at rcorbin@corbinpartners.com.

*** The authors outline practical solutions to support the reliability and validity of online surveys from a social science perspective. While illustrative legal authority is cited, a full analysis of the case law is beyond the scope of the article. The authors acknowledge with thanks the careful reviews of TMR editors.

- B. False Responses 1114
- C. Attentiveness 1114
- VI. Designing the Survey to Minimize Faking, Inattention,
and Cheating 1116
 - A. Addressing Faking..... 1116
 - B. Addressing Inattention to Instructions and
Questions 1117
 - C. Addressing Cheating..... 1119
- VII. Post survey Verification and Analysis 1120
- VIII. Conclusions 1121

I. INTRODUCTION

In many countries, including the United States and Canada, it is well established that surveys are admissible as evidence in intellectual property disputes.² Surveys were traditionally carried out by interviewers, surveying people in person or by telephone.³ Towards the end of the 20th century, technology made it possible to conduct surveys over the Internet without an interviewer. While corporations embraced Internet surveys,⁴ courts were initially more cautious about their inherent limitations.⁵ After all, the identity of the survey participants was not verifiable by sight or sound; participants' understanding of questions, on which a legal issue might turn, could not be confirmed in person; no interviewer was present to keep participants engaged. However, as with other technological breakthroughs, Internet surveys yielded new benefits that were too valuable to ignore, including cost-efficiency and wider coverage of populations than in-person surveys could ever practically achieve. Litigants are now successfully convincing courts to accept Internet surveys: the authors did not locate any decision published within the last five years rejecting Internet survey evidence solely because of its method of administration. But a definitive statement of the requirements necessary for such a survey to meet the high evidentiary standards of modern litigation is still lacking.

To meet this need, this article presents guidelines for addressing the specific vulnerabilities of Internet surveys, and proposes a consistent standard for use by courts to assess the weight to be accorded such surveys for resolution of intellectual property disputes. Most of the case law and examples in the article arise from

2. See, for example, Ruth M. Corbin & Kelly Gill, *Survey Evidence and the Law Worldwide* (2006) and Claus Eckhardt (ed.), *Trademarks 2017* (2016), the latter containing separate chapters on evidentiary practices in more than three dozen countries, and identifying countries where survey evidence is now explicitly admissible.

3. See, e.g., the many hundreds of pre-2000 cases cited in Ruth M. Corbin and A. Kelly Gill, *Survey Evidence and the Law Worldwide* (2006).

4. See Dick McCullough, *Web-based Market Research: The Dawning of a New Age*, *Marketing News*, Sept. 14, 1998, for one of the earliest discussions of using Internet surveys to guide business decisions.

5. See, e.g., *Columbia Univ. v. Columbia/HCA Healthcare Corp.*, 964 F. Supp. 733, 747 (S.D.N.Y. 1997) (Plaintiff introduced an "Internet health survey" conducted by defendant showing that four out of 1700 respondents answered yes to the question whether they had ever used a Columbia facility before and identified plaintiff's facility or a doctor associated with plaintiff as the facility; "although evidence of some confusion," the Internet survey was "entitled to little if any weight both because of the extremely small numbers involved and because there was no showing that supported the trustworthiness of the survey methodology"); see also Robert H. Thornburg, *Trademark Surveys: Development of Computer Based Survey Methods*, 4 *J. Marshall Review of Intellectual Property Law* 91, 114, 117 (2005) (noting use and acceptance of online and computer-based surveys in 2005 was not well received by the courts and estimating a lapse of six years from introduction of Internet technology to ready acceptance by courts of Internet-based survey evidence).

United States and Canada, two countries where guidance on standards for litigation surveys is thoroughly documented.⁶ However, the scientific principles on which these standards are founded are applicable to surveys worldwide. Thus, the conclusions of the article can plausibly extend to all countries where survey evidence is accepted by courts and tribunals. Applying these guidelines will help counteract the potential vulnerabilities of Internet surveys and bring the evidence in line with the scientific principles governing surveys before the Internet era. Ignoring the guidelines leaves the resulting evidence susceptible to criticism and discounting by the courts.

II. THE NEED FOR INTERNET SURVEY STANDARDS IN THE COURTS

Internet surveys are past the stage of being novel. Now widely employed by businesses and by university researchers, Internet surveys are referred to by the Insights Association⁷ as today's "most popular means of collecting data."⁸ Popularity alone may not be a solid basis for reliance on Internet surveys as evidence in court. However, consistent use of Internet surveys to support decision-making demonstrates a level of confidence in the methodology by businesses and survey professionals that provides a reasonable basis for courts to accept such surveys as expert evidence. For example, Rule 703 of the United States Federal Rules of Evidence provides that an expert may base his or her testimony on data if the data is "of a type reasonably relied upon by experts in the...field" in forming opinions or inferences.⁹ Experts need only point to the surging use of Internet surveys by companies and in other social science research to justify their own willingness to "reasonably rely" upon such surveys. Indeed, experts have successfully relied on Internet-administered surveys in a number of trademark cases.¹⁰

6. Federal courts in United States and Canada commissioned guidelines for judges' use in assessing survey evidence. See Shari Seidman Diamond, *Reference Guide on Survey Research*, in Reference Manual on Scientific Evidence 229 (2d ed. Fed. Jud. Ctr. 2000); Chapter 9, *Surveys and Other Marketplace Evidence*, by Ruth M. Corbin published in Canadian Trademark Law Benchbook (2014 ed., originally commissioned by the Federal Court of Canada and now published commercially by Carswell).

7. Insights Association establishes best practices and enforces professional standards for the American market research industry. It was formed through the merger of two organizations CASRO (founded in 1975) and MRA (founded in 1957). See <http://www.insightsassociation.org/about> (last visited July 24, 2017).

8. <http://www.insightsassociation.org/article/online-mr-quality-ignorance-bliss> (last visited July 17, 2017).

9. Fed. R. Evid. 703.

10. See, e.g., *PBM Prods., LLC v. Mead Johnson & Co.*, 639 F.3d 111, 124 (4th Cir. 2011); *1-800 Contacts, Inc. v. WhenU.com*, 309 F. Supp. 2d 467 (S.D.N.Y. 2003), *preliminary injunction reversed on other grounds and remanded*, 414 F.3d 400 (2d Cir., 2005) without addressing the admissibility of the survey evidence; *Empresa Cubana Del Tabaco v. Culbro*

Some court decisions explicitly acknowledge that a survey should not be deemed unsound merely because it was conducted online.¹¹

Notwithstanding this authority demonstrating acceptance of Internet surveys, however, courts continue to question the weight that Internet surveys should be accorded.¹² While there are many published discussions by legal researchers of advantages and disadvantages of Internet surveys compared to other methodologies,¹³ no explicit test or customized standards exist for evaluating Internet surveys in the litigation context. The question that matters most to trademark lawyers and courts is this: Can Internet surveys be trusted as a basis for reliable inferences about consumer perception?¹⁴ Commonly agreed-upon standards that can be used to answer this question are overdue.

III. RELEVANCE, RELIABILITY, AND VALIDITY

A survey's probative value may be weighed using the criteria of relevance, reliability, and validity. These criteria are endorsed by the Supreme Court of Canada as the primary framework for evaluation.¹⁵ In the United States, the Reference Manual on

Corp., 2004 U.S. Dist. LEXIS 4935, 70 U.S.P.Q.2d (BNA) 1650 (S.D.N.Y. Mar. 26, 2004), defendants' motion for a stay of the injunction pending appeal denied at *Empresa Cubana Del Tabaco v. Culbro Corp.*, 2004 U.S. Dist. LEXIS 7443 without re-evaluation of the previously admitted survey evidence; *Best Vacuum, Inc. v. Ian Design, Inc.*, 2006 WL 3486879 (N.D. Ill. Nov. 29, 2006).

11. *See, e.g.*, *Anheuser-Busch, LLC v. Innvopak Sys. Pty. Ltd.*, 115 U.S.P.Q.2d 1816 (T.T.A.B. 2015) ("While we are mindful of the limitations and advantages of various types of surveys, it cannot be said that Opposer's survey was unsound merely because it was conducted online"); *R & R Partners, Inc. v. Tovar*, 447 F. Supp. 2d 1141, 1154-55 (D. Nev. 2006) (finding no merit to defendant's argument that the "expert did not identify any journals or other articles discussing the reliability of Internet surveys" where defendants "provided no evidence to indicate an actual flaw exists in the methodology of the survey"); *T-Mobile US, Inc. v. AIO Wireless LLC*, 991 F. Supp. 2d 888, 907 (S.D. Tex. 2014) (rejecting defendant's argument that the survey should not have been conducted over the Internet).

12. *See, e.g.*, *Kraft Foods Group Brands LLC v. Cracker Barrel Old Country Store*, 735 F.3d 735, 742 (7th Cir. 2013) ("[I]t's very difficult to compare people's reactions to photographs shown to them online by a survey company to their reactions to products they are looking at in a grocery store and trying to decide whether to buy."); *His & Her Corp. v. Shake-N-Go Fashion, Inc.*, 2015 U.S. Dist. LEXIS 44874, at *31 (C.D. Cal. Mar. 30, 2015) ("the Ninth Circuit has noted that the fact that a likelihood of confusion survey 'was conducted over the internet (thereby failing to replicate real world conditions)' can detract from the weight given to a survey"), *citing* *Fortune Dynamic, Inc. v. Victoria's Secret Stores Brand Mgmt., Inc.*, 618 F.3d 1025, 1037-38 (9th Cir. 2010).

13. *See* Gabriel M. Gelb & Betsy D. Gelb, *Internet Surveys for Trademark Litigation: Ready or Not, Here They Come*, 97 TMR 1073-88 (2007) (hereinafter, "Gelb & Gelb"); Hal Poret, *Comparative Empirical Analysis of Online versus Mall and Phone Methodologies for Trademark Surveys*, 100 TMR 756-807 (2010).

14. *See, e.g.*, Shari Seidman Diamond & Jerre B. Swann, *Trademark and Deceptive Advertising Surveys: Law, Science, and Design* 308 (2012).

15. *Masterpiece Inc. v. Alavida Lifestyles Inc.*, 2011 SCC 27 (S.C.C.); *see also* Ruth Corbin, *The Moron in a Hurry: a Creature of Law or Science?*, in *Annual Review of Civil Litigation* (Toronto: Carswell, 2015), recommending these criteria as the components of a

Scientific Evidence¹⁶ enumerates questions for evaluating a survey's trustworthiness, but the questions themselves are components of the same scientific criteria of relevance, reliability, and validity. Since these three criteria are universal principles of social science, it is reasonable to expect them to be appropriate standards for evaluation in other jurisdictions as well.

Relevance concerns the extent to which a survey tends to prove the matter sought to be proved (e.g., that confusion is likely, that the mark has acquired distinctiveness, or that the mark is generic). Reliability refers to "reproducibility" of the survey result, should the survey be administered to a different sample of people in the pertinent population on a different occasion. Validity refers to the extent to which a survey measures what it sets out to measure in its stated mandate or objectives.

A. Relevance

Relevance of survey evidence will not ordinarily depend on the method by which it is collected. Indeed, relevance of any evidence is a matter to be argued by the litigator and is the prerogative of the court to decide.¹⁷ Experts presenting Internet survey evidence may have an opinion on relevance, but it is an opinion that will not necessarily be welcomed.¹⁸ Experts are expected to stick to their own areas of expertise, namely, the reliability and validity of the information on which their opinion depends. For these reasons, "relevance" is not further addressed here.

B. Reliability

Reliability, on the other hand, is a key requirement for data quality, and the way the survey was conducted will greatly affect

tripartite test for all social science evidence, building upon an earlier analysis by The Honourable Justice Todd L. Archibald and Jeremy Fox (2015), *Examining the Reliability of Expert Soft Science Evidence in the Courtroom: The Art and Science of Persuasion*, in Annual Review of Civil Litigation (Chapter IV) (Toronto: Carswell).

16. Shari Seidman Diamond, *Reference Guide on Survey Research*, in Reference Manual on Scientific Evidence 229 (2d ed. Fed. Jud. Ctr. 2000).

17. See, e.g., Fed. R. Evid. Rule 401 & Advisory Committee Notes ("Problems of relevancy call for an answer to the question whether an item of evidence, when tested by the processes of legal reasoning, possesses sufficient probative value to justify receiving it in evidence. Thus, assessment of the probative value of evidence that a person purchased a revolver shortly prior to a fatal shooting with which he is charged is a matter of analysis and reasoning").

18. Determination of relevance to the "ultimate issue" is considered to be the purview of the court. For example, in *R. v. Mohan*, [1994] 2 S.C.R. 9, the court stated: "Relevance is a threshold to be decided by a judge, as a question of law." In detailing her experience on the bench with survey evidence, retired Justice Barbara Reed of the Federal Court of Canada criticized several presumptions of a survey witness, in a case involving the PROZAC trade dress, including his rejection of the relevance of a line of questioning put to him by opposing counsel. See Justice Reed's chapter in Ruth M. Corbin & A. Kelly Gill, *Survey Evidence and the Law Worldwide* (Markham: LexisNexis, 2006).

the reliability of the results. Imagine a survey presented with the caveat that, if the survey were administered again, the reported results may or may not be reproduced. What use would it be? Because reliability depends on obtaining a random, representative sample of the population, Internet surveys are “reliability-challenged.” There is no published list or so-called sampling frame of all Internet addresses available for obtaining a true random sample of the population via the Internet. Some cohorts of the population do not even use email or the Internet. Pew Research Center reported in 2016 that more than four in ten Americans over the age of 65 do not use the Internet. Usage is also below-average among lower income households and people living in rural areas.¹⁹ So some members of the overall population have no chance at all of being captured in a random sample obtained via the Internet. Internet survey critics would therefore seem to have an instant basis for a reliability challenge, given that Internet surveys could not possibly capture all segments of a population, particularly if non-Internet users are part of the target audience. As will be discussed in the next section, the challenge is not insurmountable. Reliability can be strengthened in well-defined ways, to meet a level of “best practice.”

C. Validity

Reliability, while important, says nothing about the truthfulness of survey responses or correctness of claims about what the survey measures, which instead are measures of the survey’s validity.²⁰

Validity generally refers to whether the right elements were measured in the right way.²¹ There are different ways of

19. Monica Anderson & Andrew Perrin, “13% of Americans don’t use the Internet. Who are they?”, report of Pew Research Center, September 2016, available at <http://www.pewresearch.org/fact-tank/2016/09/07/some-americans-dont-use-the-internet-who-are-they/> (last visited Aug. 9, 2017).

20. Judges sometimes use the terms “reliability” and “validity” interchangeably, not recognizing the importance of distinguishing these scientific terms of art. See Ruth M. Corbin, *When is a Bathroom Scale Unlike a Political Poll? When it’s Reliable*, *Vue* 12, 12 (2006), for a detailed discussion of accepted distinction in science between reliability and validity, though they may on occasion be interdependent.

21. For example, in a survey of trademark confusion, the “right elements” are likely to be the perceptions of the disputed trademark features, and not the perceptions of the product overall. (That is, it is necessary to prove that the offending trademark elements are responsible for causing whatever confusion is found to exist.) Measuring the elements in the “right way” means asking about their source significance. (It would generally be insufficient, for example, to demonstrate that a certain trademark just brings to mind another owner’s product; it would need to be shown that consumers believe the two products to originate or be licensed by the same source.) An illustrative case is *Walt Disney Productions v. Fantasyland Hotel Inc.*, [1994] A.J. No. 484, 20 Alta. L.R. (3d) 146 (Alta. Q.B.), *aff’d* [1996] A.J. No. 415, 38 Alta. L.R. (3d) 441 (Alta. C.A.). In that case, the survey question about “what comes to mind” produced word associations between West Edmonton Mall’s erotic “Fantasyland Hotel” and

characterizing or confirming validity. “External validity” means that one would look to the world outside the laboratory or survey language, to confirm that the inferences hold true in everyday life. Elections, for example, allow one to validate the results of a pre-election poll. For trademark matters, actual instances of confusion for reasons similar to those discovered in the survey provide external validity of the survey results. “Convergent validity” refers to the congruency of measurements from different methodologies. For trademark matters, findings in telephone surveys or mystery shopping studies²² similar to results obtained in an Internet survey would provide convergent validity.²³ “Face validity” means that the answers to the questions themselves provide *de facto* evidence. Answers to open-ended, undirected questions (“Any comments on the product you just saw?”) that reveal confusion have face validity. Each category of validity presents an opportunity for an expert to demonstrate an aspect of scientific integrity.

Validity is mainly a function of question wording and sequencing, and the extent to which the questions elicit truthful or mindful answers. There is no reason that Internet surveys cannot be as valid as surveys done by other methods. In fact, Internet surveys have features that can enhance validity. For example, without an interviewer present, participants may feel less inhibited when answering sensitive questions.²⁴ Furthermore, computer programming can produce customized question wording that automatically incorporates answers previously given by a respondent verbatim.²⁵ Automatic programming is much superior to expecting live interviewers to go back to previous questions to interpret and extract what a respondent said earlier, or to follow complex instructions about which questions to skip to, depending on

Disney’s Fantasyland amusement park; the court found the evidence insufficient to sustain a finding of confusion.

22. Mystery shopping entails collection of first-hand information about the customer experience by researchers presenting themselves as customers. Trained researchers visit stores or other retail organizations, send Internet inquiries, or phone for service, posing as customers of the establishments they are contacting. See Ruth M. Corbin & Sarah Carnegie, *Mystery Shopping Raised to Scientific Evidence*, *Vue*, September (2009).

23. See Ruth M. Corbin & Fiona Isaacson, *Surveys on a Tightrope—the Convergent Validity Net*, 24 *Intell. Prop. J.* 265-78 (2012).

24. Paul J. Lavrakas & Jeffrey A. Stec, *Survey Research in Litigation*, in *Litigation Services Handbook*, 2016 Cumulative Supplement (Roman L. Well, Daniel G. Lentz & Elizabeth A. Evans eds., 5th ed. 2016), at p. 36.

25. Gelb & Gelb, *supra* note 13, at p. 1077. An Internet survey can be programmed in such a way that it “presents initial questions, and then displays succeeding questions based in part on answers a respondent has given previously (a procedure known as ‘branching.’) Often, the questions displayed include a word or phrase provided by the respondent in a previous answer (known as ‘piping.’)”

a respondent's earlier answers.²⁶ The following examples of wording customization are adapted from fact scenarios of published cases, that used in-person surveys; the instructions in capital letters illustrate the opportunity for a seamless pre-programmed customization:

- “*You said you thought French Press was a [INSERT ‘BRAND’ or ‘DESCRIPTIVE NAME’ DEPENDING ON PREVIOUS ANSWER]. Why do you say that?*”²⁷
- “*Why do you believe this building block is made by [INSERT COMPANY NAME USED IN PREVIOUS ANSWER]*”²⁸

This type of customization personalizes the survey experience, thereby building rapport and facilitating truthful answers.

The main risks to the validity of Internet surveys are that anonymous respondents may consult other people or sources for answers to questions, or may be inattentive or unmotivated to answer thoughtfully. If an Internet survey, or any survey, is unable to elicit true attitudes or beliefs from respondents, then it cannot produce valid evidence of trademark perception.

As shown above, criticisms about reliability and validity can be isolated to specific elements of an Internet survey. Quality control measures are available to deal with these issues, and to strengthen reliability and validity of litigation surveys to meet a level of best practice. The quality control measures are explained in the next four sections.

IV. PANEL SELECTION AND RELIABILITY— ISSUES AND REMEDIES

Internet surveys conducted for litigation are usually administered through panels consisting of very large numbers of previously recruited volunteers who answer repeated surveys over the course of their panel membership.²⁹

26. This modification of question wording in real time can also be achieved by a human interviewer who asks questions from an Internet-connected tablet. Even for such surveys, the presence of a human interviewer may be superfluous.

27. *Bodum USA, Inc. v. Meyer Housewares Canada, Inc.*, 2013 FCA 240.

28. Battles were held in Canada, the United States and Britain over the trade dress of MEGA BLOKS, alleged to be confusing with that of LEGO. Canada's Supreme Court decided the matter once and for all in *Kirkbi AG v. Ritvik Holdings Inc.*, [2005] 3 SCR 302, 2005 SCC 65 (CanLII), with an outcome similar to that in other countries, namely, that LEGO could not protect as trademarks the functional aspects of its products.

29. Panels include hundreds of thousands to several million members. See Reg Baker et al., *AAPOR Report On Online Panels*, 74 The Pub. Opinion Q., 711-781 (2010). Industry guidelines do not discourage panel companies from retaining members as long as members stay active. See the website of recognized world survey association ESOMAR, addressing questions about Internet panels, at https://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR_26-Questions-To-Help-Research-Buyers-Of-Online-Samples.pdf (last visited July 23, 2017). The number of surveys that any one

Such panels evolved out of necessity. Sampling the population randomly by Internet is not feasible for many reasons: First, there is no universal directory of email addresses. Even if there were, it would be impossible to know the extent of duplicate addresses for the same person, necessary to make claims of superior reliability. Second, privacy laws in certain jurisdictions³⁰ make it unlawful, or potentially unlawful, to send mass solicitations to private email addresses. Third, soliciting survey participation through clickable hyperlinks on websites is not likely to lure potential respondents to interrupt their Internet browsing to provide anonymous responses on remote topics for no compensation. Indeed, survey solicitation on sponsored Internet sites produces what are referred to deprecatingly as strictly “entertainment surveys”³¹—hardly a respectable foundation for litigation evidence.

For these reasons, volunteer Internet survey panels were created and are now offered by hundreds of commercial suppliers worldwide.³² Panels vary in size, some exceeding a million people. Panel members are asked for demographic information and details about their use of various products and services. This information is maintained in the panel database. For any given survey request, a panel manager can recruit panel members satisfying criteria of interest to a survey client. Some courts accept surveys wholly managed not by the testifying expert witness, but by third party panel administrators,³³ seeming to stray farther from the hearsay rule of evidence than the normal leeway allowed for experts.³⁴

individual could be invited to complete varies across panels. According to a 2012 study by Grey Matter Research, invitations to single panel members could exceed fifty per month, though the member might only answer or qualify for a small number of those. The study entitled “More Dirty Little Secrets of Online Panel Research” is available upon request at http://www.greymatterresearch.com/index_files/Online_Panels_2012.htm (last visited Oct. 29, 2017).

30. The nuances of different countries’ privacy laws are beyond the scope of this article. By way of example, Canada’s “anti-spam legislation” (CASL) explicitly prohibits mass emailing without consent of each recipient. See <http://crtc.gc.ca/eng/internet/anti.htm> for more detail on this legislation.

31. See Gelb & Gelb, *supra* note 14, at 1080.

32. ESOMAR defines panels as databases of potential research participants who declare that they will cooperate with future data collection requests if selected. (See <https://www.esomar.org/what-we-do/about-us>). Its guidelines promote transparency with clients and provides ways to ensure online sample quality.

33. In *In re NJOY, Inc. Consumer Class Action Litigation*, 120 F. Supp. 3d 1050, 1080-81 (C.D. Cal. 2015), the defendant attacked the plaintiffs’ Internet survey on the basis that “although designed and analyzed by [plaintiffs’ expert, it] was carried out by a third party survey company, Qualtrics.com.” The court rejected this argument, noting that defendant “cites no case law holding that surveys conducted by third parties are inherently ‘flawed’ or unreliable, and the court has located none. What little case law there is on the subject in fact suggests there is nothing inherently problematic with having a survey conducted by a third party.”

34. See, e.g., Ruth Corbin, R.S. Jolliffe, & A.K. Gill, *Trial by Survey*, (2000) in which the authors give a historical review of the Expert Opinion exception to the Hearsay Rule; the

A. Problems with Internet Panels

1. Identity

Internet panels raise questions of reliability: do we know who these Internet survey participants really are? Participants may or may not be completely honest at sign-up time, particularly if they are privacy-sensitive or if they worry about spam-email that might take advantage of their demographic information. Verifying information at the sign-up stage is the responsibility of a good panel company. But not all panel companies are diligent in this respect. In one passing off case regarding impression of car company slogans,³⁵ a survey expert testified about the quality control used by the panel company to verify the identity of participants in the Internet panel [referred to for this case example as “PanelCo.”] The name, address, and telephone details of panel participants were reviewed at PanelCo’s sign-up stage, the expert reported, to validate participant authenticity at least at face value. An inspection was further done by PanelCo, the expert claimed, to verify whether the name was credible (he gave examples of names of famous people that would be rejected), whether the street address actually existed, and whether the postal codes matched the address location. Opposing counsel doubted the expert’s assurances of quality control, and engaged individuals to sign up with PanelCo, using readily detectable false information. Opposing counsel’s purpose was to see if falsehoods would be caught, as the expert had assured they would be. One of the hired individuals signed up under the name “Napoleon Bonaparte.” Another used a street name that did not exist in the given city. Another used a postal code inconsistent with the city. All individuals—having submitted easily checkable false information—were accepted as panel members, and were invited to participate in PanelCo surveys. This set of revelations was documented in an opposing expert’s affidavit. The case settled soon thereafter.

2. Representativeness

Verifying respondent identities is not the only challenge. Representativeness of a given single panel sample has also been challenged in the courts.³⁶ Can any given panel be representative

authors observe that courts still expect that experts will personally oversee the data collection process, in order to give the court first-hand confirmation of their confidence in the data underlying their opinion.

35. *Bennett Dunlop Ford Sales Ltd. v. Kia Canada Inc.*, Federal Court of Canada File No: T- 1993-10, settled out of court after exchange of evidence, in 2014.

36. See *Parallel Networks Licensing, LLC v. Microsoft Corporation*, Civil Action No. 13-2073 (KAJ), 2017 U.S. Dist. LEXIS 29613, at *9-10 (D. Del. Feb. 22, 2017) (“According to Dr. Isaacson, survey respondents were recruited through an internet survey panel organized by Survey Sampling International (‘SSI’). ... But Dr. Isaacson does not provide any details on

of the population to be tested? If the same survey were run on two different panels, would similar results be obtained? Empirical investigation has yielded the disturbing finding that an identically worded survey, administered to participants with similar demographic profile on two different commercial panels, may produce markedly different results.³⁷ Factors proposed as explanations for such variations across panels include the following:

- different contexts in which the survey is presented, or different visual layouts;
- different respondent motivation, arising from amounts of compensation offered by different panels; and
- different recruiting sources for the participants;
- different expectations or frames of thinking based on the wording of the “appeals” at the time of recruiting.

Surveys that blend members from more than one panel may be no better, because the variance, or “irrelevant noise” factor, may be significantly increased.³⁸ The point to be made is that differences between Internet panel populations may be as large as differences among populations of two provinces or states. The people in the different panels may “seem” the same, but were pre-conditioned by very different environments, which may result in distinctly different attitudes on a given issue.³⁹ High variance in a data set makes it harder to detect an effect that exists in the overall population (referred to in science as a “Type II error”⁴⁰). Consider a likelihood of confusion case, wherein the plaintiff’s single panel-based Internet

the particular panel he used for the survey. It is not clear if the panel drew from a representative sample of the United States population, of IT professionals, or from some other group....Without additional information, it is impossible to know whether the participants of the underlying panel deviate in some meaningful way from the target population. For example, it could be the case that some of Microsoft’s customers are frugal and wish to save money—motivating them to hire less experienced IT professionals and to optimize their network configuration based on cost—while other customers are more interested in network performance—motivating them to hire experienced IT professionals and to configure their networks to minimize service times. If the SSI panel, used for the survey, screened individuals based on income, then the survey results could systematically over represent one set of Microsoft customers while under representing another”.)

37. See Jon Puleston & Mitch Eggers, *Dimensions of Online Survey Data Quality. What Really Matters?*, Proc. of ESOMAR Congress (2012).

38. See ESOMAR, *28 Questions to Help Buyers of Online Samples* (2012), <https://www.esomar.org/uploads/public/knowledge-and-standards/documents/ESOMAR-28-Questions-to-Help-Buyers-of-Online-Samples-September-2012.pdf> (last visited Aug. 9, 2017) (hereinafter, “ESOMAR 28”).

39. See R. Walker, R. Pettit, & J. Rubinson, *The Foundations of Online Research Quality—Executive Summary 3: Inter Study Comparability and Benchmark Analysis*, New York: Advertising Research Foundation (2009).

40. Statistically, a Type II error means a failure to reject a false null hypothesis. Put differently, it means the inability to detect a causal effect in a survey sample when in reality the effect does exist in the universe. This is otherwise known as a “false negative.” A Type I error is a false positive, or the detection of an effect that is not actually present.

survey finds confusion between a junior and senior brand. The defendant proffers a blended-panel Internet survey (using two entirely different panels) that shows the opposite: no confusion. If the defendant's survey has exaggerated variance, it may be the case that confusion exists but was not detected. In such a case, the opposite results shown by plaintiff's and defendant's surveys may be the result of different Internet panel selection, as opposed to (or perhaps in addition to) other differences in the stimuli or questioning.

Would a larger sample size selected from any panel not help control unwanted noise and increase representativeness? While it is statistically true that a large sample size reduces the influence of noise in detecting "true" effects, a large sample size is no panacea for poor representativeness. A well-known example comes from the 1936 US presidential election survey where *Literary Digest* surveyed nearly two and a half million people (a very large sample for the time), and predicted that Alfred Landon would win the election against Franklin D. Roosevelt. The survey was wrong. The tested population consisted of people listed in directories of citizens of upper income groups, whose views were not those of the majority.

3. Motives of Volunteers

Another concern expressed about Internet panels is that they consist largely of volunteers, who want to be part of the panel club, and are willing to answer several surveys a year in return for compensation or entries in contests. They are "motivated survey-takers," for better or worse. Over-motivated survey respondents present cause for concern, whether the survey is conducted in person, by telephone, or over the Internet, because such respondents may have little or no investment in the truthfulness of their answers to any particular survey.

The issue arose in a high-profile pharmaceutical trademark case in Canada, concerning the PROZAC drug.⁴¹ A mall survey submitted by the plaintiff was entirely discounted by the court for its insufficient quality controls. Among the identified weaknesses was the survey company's failure to screen out people who "hung out" in malls where survey companies operated, waiting to be paid for taking surveys, and communicating with each other about answers to the qualifying questions for that day's survey. Justice Reed in that case famously coined the term "Mallies" to describe such people. Most Canadian trademark surveys since that case

41. *Eli Lilly and Company v. Novopharm Ltd.*, 1997 CanLII 5008 (FC), *aff'd* [2001] 2 FC 502; Justice Barbara Reed, who presided over the dispute, subsequently described the survey evidence at trial in a chapter published in Ruth M. Corbin & A. Kelly Gill, *Survey Evidence and the Law Worldwide: A Reference Text for Lawyers, Jurists and Social Scientists* (2006).

feature a qualifying question to screen out Mallies.⁴² Internet panels arguably contain a similar ingredient of motivated volunteerism that needs to be held in check.⁴³

Where regulations for panel recruitment exist, they are designed principally to protect consumers from unwittingly giving up personal information without realizing that their personal information would be kept in the permanent records of a panel company.

International Organization for Standardization (“ISO”)⁴⁴ regulation 26362:2009, embraced by many reputable panel providers, recommends a double opt-in measure for panel volunteers, by obtaining their consent at two separate points of the recruitment process. While protecting against panel companies’ abuse of consumer naiveté, the ISO standard surely reduces the variability of personality types that end up on panels, that is, by limiting them to people willing to invest time to pause, consider and agree to be a continuing volunteer. Again, one is reminded of some of the foregone advantages of earlier survey methods. Random telephone surveys, for example, were once lauded for their ability to capture people who would not ordinarily volunteer for surveys, but who could be persuaded on individual occasions to answer questions from a courteous interviewer at the other end of their phone line. The ISO standard for Internet panels reduces the chance of capturing less-willing respondents, thereby risking a loss of representativeness. This paradox needs to be addressed.

B. Risk Management of Internet Panels

We have articulated thus far recognized risks of Internet surveys with respect to the assembly of random samples, risks not present or not as significant with older-style methods. These risks can be managed; the first step in risk management is careful selection of a panel company. Panel companies suitable for

42. A typical wording is as follows: “Within the past 30 days, have you or have you not participated in a survey conducted in a mall?”

43. To the authors’ knowledge, this risk of distortion of results by “professional respondents” in Internet surveys has not been raised in adversarial legal argument. The recognized international standards organization for the survey industry, ESOMAR, does not set any guideline to address the risk identified here.

44. ISO is an independent, non-governmental international organization that brings together experts to develop voluntary, consensus-based, market relevant international standards. See <https://www.iso.org/about-us.html> (last visited July 17, 2017). Courts define ISO similarly. See *Thatcher v. Trans World Airlines*, 69 S.W.3d 533 (Mo. Ct. App. 2002) (“ANSI (American National Standards Institute), formerly known as ASA, is a voluntary membership organization that develops consensus standards nationally for a wide variety of devices and procedures. ISO is the international standards organization that performs a similar function”). The relevance of ISO standards is acknowledged in case law in other areas, such as products liability. See, e.g., *Holst v. KCI Konecranes Int’l Corp.*, 699 S.E.2d 715, 390 S.C. 29 (Ct. App. 2010).

supplying data to litigation will have invested in the following quality controls, that expert witnesses will be expected to disclose:⁴⁵

- Recruitment procedures that verify respondent identity and demographic information;
- Recruitment procedures that reach out to people who may not actively volunteer; and
- Neutral wording of recruitment invitations to avoid biasing the participants' expectations, or understating the importance of thoughtful participation.

The international research association, the European Society for Opinion and Market Research ("ESOMAR"), suggests questions to be asked of potential online sample providers⁴⁶ to permit clients to at least assess the fit between what they need and what a panel company offers. The standards verified through asking such questions can be documented in an expert report, and used to compare the credibility of competing surveys submitted to a court. ESOMAR's recommended questions are enumerated in summary form below, accompanied by case references where courts have expressed explicit interest in the answers to just such questions.

- What is the source of the panel members?
- Are there checks on respondents' identity? In *3M v. Mohan*,⁴⁷ an online survey of doctors and nurses was given greater weight because the panel provider collected a substantial amount of identifying information about these medical professionals at the time of recruiting them onto the panel. Also, for purposes of sending out the monetary honoraria upon survey completion, the panel provider collected the mailing addresses of survey participants, thereby further verifying their identities.

45. Lack of clarity on how participants to a survey were solicited has attracted notice of courts. For instance, in *Thermolife International, Llc v. Gaspari Nutrition, Inc.*, No. CV-11-01056-PHX-NVW, 2014 U.S. Dist. LEXIS 3426, at *6 (D. Ariz. Jan. 10, 2014), the court noted: "Mr. Berger's report does not state when the survey was conducted or how survey participants were solicited," and excluded the survey. On appeal, however, the Ninth Circuit found that the district court's exclusion of Internet survey evidence was clearly erroneous because "[a]lthough the district court faulted the survey's biased questions and unrepresentative sample, neither defect was so serious as to preclude the survey's admissibility." 648 Fed. Appx. 609, 614 (9th Cir. 2016), citing *Fortune Dynamic*, 618 F.3d at 1037-38.

46. ESOMAR is a non-profit organization that plays a significant role in market research industry by setting quality standards as well providing advocacy support. It has existed since 1947 and has 5000 individual and 500 corporate members. ESOMAR's advice on panel assessment is contained in ESOMAR, *28 Questions to Help Buyers of Online Samples*, at <https://www.esomar.org/uploads/public/knowledge-and-standards/documents/ESOMAR-28-Questions-to-Help-Buyers-of-Online-Samples-September-2012.pdf> (last visited Aug. 14, 2017).

47. *3M Co. v. Mohan*, Civ. No. 09-1413 ADM/FLN, 2010 U.S. Dist. LEXIS 124672 (D. Minn. Nov. 24, 2010), *aff'd* No. 2011-1328, 2012 U.S. App. LEXIS 10753 (Fed. Cir. May 29, 2012).

- Will it be possible to compare results of panel members based on the source from which they were recruited? This is called split-sample reliability testing,⁴⁸ a quality control step long-recognized in the social science literature.
- Is the panel company ISO certified?⁴⁹
- What is done to ensure no duplicate responding? If more than one panel is used, the researcher needs the cooperation of the panel providers to avoid duplication. Knowledge of the blending process also provides an insight into how de-duplication of respondents across sources is handled.
- Are ethnic minorities or other hard to reach groups, if pertinent to the dispute, captured in the panel in suitable proportion? Certain ethnic minorities have a distinctly higher proportion of mobile-only Internet access,⁵⁰ and many Internet surveys are designed to only be used from computer screens as opposed to on mobile devices.⁵¹
- What is the nature of incentives offered for participation? Are incentives likely to encourage speeding, inattention, or dishonesty? The court in a class action case criticized the absence of clarity regarding incentives or compensation to prospective survey respondents.⁵²

ESOMAR's advice to put such questions to prospective panel suppliers is not itself a guarantee of high-quality data. Nevertheless, asking such questions, and accounting for answers, is a demonstration of desirable due diligence.

48. See Michael Furr, *Split Half Reliability in Encyclopedia of Research Design* (Neil Salkind ed., 2010), <http://methods.sagepub.com/reference/encyc-of-research-design/n431.xml> (last visited Oct. 15, 2016).

49. Access to ISO's certification information is available at http://www.iso.org/iso/catalogue_detail?csnumber=43521 (last visited Aug. 9, 2017).

50. See, e.g., Monica Anderson, *Racial and Ethnic Differences in How People Use Mobile Technology*, Pew Research Center (2015), <http://www.pewresearch.org/fact-tank/2015/04/30/racial-and-ethnic-differences-in-how-people-use-mobile-technology/> (last visited Aug. 9, 2017). This Pew Research Center report shows that in US "some 13% of Hispanics and 12% of blacks are smartphone-dependent, meaning they don't have a broadband connection at home and have few options for going online other than their cellphone. In comparison, only 4% of white smartphone owners rely heavily on their cellphone for online access." This disparity demonstrates the importance of incorporating a mobile-access component in online surveys.

51. On the other hand, where response by mobile devices is permitted, a concern has been raised about possibly lower quality of response. See Aigul Mavletova, *Data quality in PC and mobile web surveys*, 31 Soc. Sci. Computer Rev. 725-743 (2013).

52. *In Re Front Loading Washing Machine Class Action Litigation*, Civil Action No. 08-51 (FSH), 2013 U.S. Dist. LEXIS 96070, at *22 (D.N.J. July 10, 2013). An Internet-based consumer survey was conducted to determine the incidence of problems and the overall satisfaction or dissatisfaction among the LG washing machine owners. The court noted: "It is unclear whether the people who took the survey were paid or otherwise rewarded to take it," and excluded the survey under *Daubert*.

V. PARTICIPANT RESPONSE AND SURVEY VALIDITY

Participants are the ultimate determiners of the quality of data they provide. They are the hidden witnesses, on whose responses an expert relies in presenting an expert opinion to the court. This section enumerates issues arising from participants themselves, and identifies best practices to address these issues.

A. Duplicates

An assumption underlying any survey is that a participant completes the survey only once. Can a researcher be certain that each respondent is unique? Perhaps not. Data indicate that 45% of U.S. panel respondents are members of more than 5 panels.⁵³ This multiple participation becomes a problem when more than one panel is used to obtain “hard-to-reach” respondents—such respondents may then happily complete the same survey more than once, for multiple rewards.

Multiple responses by the same individual create particular threats to defendants in confusion cases. Plaintiff surveys are typically treated as probative of material confusion if they produce evidence at least in the range of 10% to 20% confusion.⁵⁴ Every time a confused respondent participates more than once, he or she brings the evidence closer to the minimum. Respondents may also sign up for the same panel under different email addresses, with false demographics. Falsely given demographics prevent a researcher from having a correctly balanced demographic sample, which in turn undermines statistical reliability. It may also undermine relevance, if the false information is pertinent to the trademark dispute. For example, if the audience sought for a toy trademark study is mothers of young children, then any panel participants falsely reporting on ages of their children, or their own gender, could be wrongly captured in the survey sample. Panel provider websites describe various tactics for verifying identity, including digital fingerprinting that records the IP addresses of respondents.⁵⁵ But there is no uniform industry practice agreement.

53. See Online Panel Research: A Data Quality Perspective 43-44 (Mario Callegaro, Reginald P. Baker, Jelke Bethlehem, Anja S. Göritz, Jon A. Krosnick, & Paul J. Lavrakas, eds., 2014).

54. Gerald Ford, *Survey Percentages in Lanham Act Matters*, in *Trademark and Deceptive Advertising Surveys* 314 (Shari Diamond & Jerre B. Swann eds. 2012) (“While it is true that some courts have not accorded probative weight to survey results between 10 percent and 20 percent, numerous courts have found likelihood of confusion surveys with percentage results between 10 percent and 20 percent probative of a likelihood of confusion”).

55. Digital Fingerprint (or Device ID or Machine fingerprints or Machine ID) is defined by ESOMAR as “A technology-enabled system that establishes a set of configuration data about a research participant’s device (computer, smartphone, etc.), which can be used to create a machine or device fingerprint. Such systems assume the “machine fingerprint” uniquely identifies a device using settings and characteristics associated with an individual

B. False Responses

Respondents may falsely report product usage or job responsibilities if they believe it would qualify them for a survey reward. This is problematic for trademark surveys, which depend on particular pertinent populations—namely those people who make purchases or influence purchases in the trademarked product/service category. Courts are aware of this vulnerability of Internet surveys. In *Dwyer Instruments, Inc. v. Sensoccon, Inc.*,⁵⁶ the court noted the need for the expert to address whether respondents sitting at the computer were actually part of the identified universe. In *First National Bank of Omaha, Inc. v. MasterCard International*,⁵⁷ the court noted: “The fact that respondents were required to take time to complete the internet survey in exchange for a small amount of money, moreover, suggests the possibility that the individuals who volunteered may not have been the professionals who make high-level decisions concerning the banks’ adoption of smart card programs. It is unclear whether, and to what extent, the survey reached the decision-makers whose potential confusion is relevant to this litigation.” In *National Financial Partners Corp. v. Paycom Software, Inc.*,⁵⁸ the respondents to the online survey were those with job responsibilities in payroll, insurance, and personnel. The court here noted: “Of thirty-one survey participants contacted by an independent firm, ten were not actually qualified to take the survey. This calls into question how many of the other 201 participants were qualified.”⁵⁹

C. Attentiveness

Valid results depend on respondents’ reading the instructions, reading each question carefully, and answering thoughtfully. Alas,

device or, potentially, an individual user account.” See https://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR-GRBN_Online-Sample-Quality-Guideline_February-2015.pdf (last visited July 17, 2017). One example of many that accomplishes digital fingerprinting is the OPTIMUSID technology designed by social media sampling firm, Peanut Labs. Peanut Labs has removed the webpage that described their OptimusID technology. However a news release that provides the same information is available at <http://www.marketwired.com/press-release/peanut-labs-unveils-optimusid-to-resolve-online-market-research-data-quality-crisis-835622.htm>.

56. *Dwyer Instruments, Inc. v. Sensoccon, Inc.*, Cause No. 3: 09-CV-10-TLS, 2012 U.S. Dist. LEXIS 21308 (N.D. Ind. Feb. 21, 2012).

57. *First Nat’l Bank of Omaha, Inc. v. MasterCard Int’l Inc.*, 03 CIV. 707 (DLC), 2004 U.S. Dist. LEXIS 13162 (S.D.N.Y. July 15, 2004).

58. *Nat’l Fin. Partners Corp. v. Paycom Software, Inc.*, No. 14 C 7424, 2015 U.S. Dist. LEXIS 74700 (N.D. Ill. June 10, 2015).

59. Checking on the integrity of the panel data base is a creative opportunity for an opposing firm, to expose one of the potential vulnerabilities of an Internet survey. When the panel is managed by a firm with rigorously implemented quality controls, the tactic is less likely to bear fruit.

that is not always the case. Reported studies have found that as many as 46% of respondents don't read instructions carefully and 11% don't read questions.⁶⁰ Respondents who are part of multiple panels may be particularly disinclined to read surveys carefully because they participate in so many surveys and either want to get through the each survey as quickly as possible or assume they understand the questions based on past experience.⁶¹ Recent evidence shows that data contamination due to respondents' careless and inattentive responding may lead to identifiable Type I and Type II errors,⁶² that is, finding an artificial effect such as confusion that may not actually exist, or alternatively failing to find a true effect such as confusion that does exist. In either case, validity is compromised: the survey is not measuring what it was designed to measure.

Ensuring that respondents read the instructions is critical for tests of genericism in the *Teflon* survey format.⁶³ A *Teflon* survey is distinctive for its intended purpose, and not a format with which most consumers would be familiar (unless they are routine respondents to surveys). In the *Teflon* survey format, participants need to first be instructed on the difference between the terms "brand name" and "common name." After being given one or more examples, they are then asked to distinguish each name on a list of names as being either a brand name or a common name. A *Teflon* survey can simply not be valid unless responders attend to the instructions.⁶⁴

60. See Daniel M. Oppenheimer, Tom Meyvis & Nicolas Davidenko, *Instructional manipulation checks: Detecting satisficing to increase statistical power*, 45 J. Experimental Soc. Psychol., 867-872 (2009) (hereinafter, "Oppenheimer et al.").

61. The average number of survey invitations to members of a panel in a month could be as low as 6.5 to as high as 52. See "More Dirty Little Secrets of Online Panel Research" (2012), <http://www.greenbookblog.org/2012/01/30/more-dirty-little-secrets-of-online-panel-research-2/> (last visited Aug. 9, 2017). It is thus possible for a panel member to receive, on average, more than one survey invitation every day from a given panel. Moreover, 45% of panel members are participants in more than 5 panels and these members of multiple panels will presumably have even more opportunities to participate in surveys.

62. Jason L Huang, Mengqiao Liu & Nathan A, Bowling, *Insufficient Effort Responding: Examining an Insidious Confound in Survey Data*, 100 J. Applied Psychol. 828 (2015).

63. See, e.g., Gerald Ford, "Intellectual Property Surveys. Annual Cumulative Update 2010," prepared for the International Trademark Association (INTA), and posted at <http://www.inta.org/TrademarkBasics/FactSheets/Documents/INTALanhamActSurveysCumulativeUpdate2010.pdf> (last visited August 9, 2017). Generally speaking, Teflon surveys require participants to categorize a given name as a brand name or common name, in order to determine whether the primary significance of the name to the consuming public is that of a brand or a generic term.

64. See, e.g., *Sheetz of Delaware, Inc. v. Doctor's Associates Inc.*, 108 U.S.P.Q.2d 1341 (T.T.A.B. 2013) (Teflon survey done without conducting a mini-test to confirm respondents understood the instructions discounted for this and other reasons; "[a]sking a respondent whether he or she understood the difference [between a brand name and a common name] is not the same as testing whether she or he understood the difference.").

Answering thoughtfully is also necessary to the validity of a survey in the *Squirt Test*⁶⁵ format, a trademark survey format involving a closed-ended question about the sources of two products, with answer options inviting the respondent to distinguish between the “same company” or a “different company.” Answering thoughtfully continues to be necessary in answering the open-ended question that follows, a question like “Why do you say that?” Respondents are free to type in as little or as much as they wish. Overly speedy participants who don’t want to “think” too much may rob the survey-taker of potentially valuable information about their perceptions.

On the other hand, excessive time taken by respondents may be indicative of a different kind of problem. For example, in response to a survey in the *Eveready Test*⁶⁶ format, where the question presented is “Who do you think puts out this product?”, respondents who want to be “right” (as research shows many respondents like to do) may try to look up the answer on the Internet. Or they may ask a family member. Or, if the option exists in the survey formatting, they may try to go back to change earlier responses to demonstrate consistency with what they plan to type next.

* * *

The issues raised in this section are about the individuals behind the numbers. Internet surveys achieve cost-effective reach of thousands of such individuals, without concern for their unique personal perspectives, motivations or answering styles. Yet the latter can introduce biasing effects. The next section explains principles of quality control that can be applied to control potential biases in three identified categories of risk: “faking,” “inattention,” and “cheating.”

VI. DESIGNING THE SURVEY TO MINIMIZE FAKING, INATTENTION, AND CHEATING

Internet surveys can be designed so as to minimize respondent behaviors such as “faking answers,” “inattention,” or “cheating.”

A. Addressing Faking

Faking occurs when respondents deliberately provide misleading or inaccurate responses to survey questions. As earlier discussed, one reason for faking may be to try to qualify for a survey

65. The Squirt test is named after for format set out in *Squirtco v. Seven-Up Co.*, 628 F.2d 1086, 1089 n.4 (8th Cir. 1980), in which a survey participant views allegedly confusing products side by side or in short succession.

66. The Eveready test was first set out in *Union Carbide Corp. v. Ever-Ready, Inc.*, 531 F.2d 366, 385-88 (7th Cir. 1976), and entails asking about the source of a product without display of any other product with which it may be allegedly confused.

without actually having the necessary pre-requisites—perhaps by falsely claiming to have bought a product relevant to the survey topic. Another reason for faking may be to appear to act consistently with social norms.

The problem of faked qualifications can be reduced by screening-questions that either identify those who are willing to give false answers, or make it difficult for “fakers” to guess how to answer qualifying questions to gain entry into the survey. One way of doing the former is to include questions in the survey to which affirmative responses are highly unlikely. For example, respondents who answer affirmatively to the following three questions are unlikely to be consistently telling the truth: *Did you purchase scuba gear in the last 3 months? Did you go snow-skiing in the last 3 months? Did you visit a fly fishing store in the last 3 months?* If the additional question “Did you buy a camera in the past 3 months?” is the actual qualifying question, then the earlier questions both help to detect likely liars, as well as make it very difficult for the respondent to guess which question is the true qualifier.

Detecting faked identification (for example, a child answering on her parents’ computer) may appear more difficult at first glance, because of respondents’ presumption that their answers will be anonymous. The authors ask each respondent to provide a name and telephone number and convenient time to be reached during a day, to allow a survey administrator to verify that it was the respondent who filled out the survey. Participants are still assured that their identification details will not be tied to their answers, and will be used for no other purpose than to verify their participation. Respondents unwilling to provide such information are prohibited from continuing.

B. Addressing Inattention to Instructions and Questions

The second category of respondent issues that can be mitigated is that of inattention, either to essential instructions or to the content of the questions themselves. Perhaps a respondent wants to get the survey over quickly, increasing his “pay per minute” for participating. Or, respondents may believe that they recognize the format from previous surveys (like the *Teflon* survey format for genericness surveys) and may feel they don’t need to pay as close attention to the instructions. Or, fatigue or distraction may lead them to answer questions quickly and without thought—even checking off boxes without reading the questions.

We first consider the need to ensure attention to survey instructions. One technique for doing so is the use of an instructional manipulation check (“IMC”).⁶⁷ An IMC works

67. See Oppenheimer et al., *supra* note 60.

approximately as follows. An initial screen page presents the instructions in the form of a title, a paragraph, a question with multiple choice response buttons, and a large “continue” button. Midway through the paragraph, respondents are instructed to ignore the multiple-choice buttons, avoid using the “continue” button, and simply click on the title to move to the next screen. Respondents who fail the manipulation check by doing one of the things they were instructed *not* to do are prompted that they are not paying enough attention and are taken back to the screen again till they get it right. Research has shown that such prompting to pay more attention does improve their answers, making their answers indistinguishable from the answers of those who “passed” the IMC initially. Without such prompting, the answers from those who initially failed the IMC digress from those who passed it.⁶⁸ A *Teflon* survey can readily incorporate such a question by first explaining the difference between the terms “brand name” and “common name,” giving examples, and then testing respondents’ understanding with questions where the answer should be obvious—like distinguishing whether “Bacardi” is a brand name or a common name, or whether “milk” is a brand name or common name. Respondents who fail to get the concept after one or two tries can be eliminated from progressing through the survey. Such a pre-quiz to *Teflon* surveys is integral to its use.⁶⁹

There is also the risk of inattention to the questions themselves. Careless responding can be detected through special sets of questions; respondents who then fall below a minimum “attention score” are terminated from the survey.

One such set of questions are “inconsistency items.” Pairs of questions can be included in the survey, relatively distant from each other, that request the same information with different wording. For example, asking for agreement or disagreement to the following two items should yield the same response if participants are paying attention: “I am an active person.” “I live an active lifestyle.”⁷⁰ Another question type to detect inattention is the “bogus item.” Certain common-sense questions would be answered in a predictable way by respondents who are paying attention. Different answers would be a sign of inattention. Examples of such items include: “I was born on February 30th [check off yes/no].” “I have

68. See p. 870, Oppenheimer et al., *supra* note 60.

69. See, e.g., Ruth M. Corbin, *Surveys and Other Marketplace Evidence*, in *Canadian Litigation*, in *Canadian Trade-mark Law Benchbook*, (D. Cameron ed., 2015); Jacob Jacoby & Lynda Zadra-Symes, *Legal Issues That Can Be Examined via Surveys*, in *Trademark Surveys*, Volume I: Designing, Implementing and Evaluating Surveys (2013); see also a full reproduction of the wording of a *Teflon* survey in *In re Calloway Golf Company* [2001] TTAB, USPTO, Serial No. 75/389,003.

70. Some work uses 7-point scale items to test inconsistency. A dichotomized (yes/no) version is simpler to manage and analyze.

been to every country in this world [check off yes/no].” Another type is a “directed-answer question.”⁷¹ Similar to the IMC discussed earlier, a directed question tells a respondent which specific response option to choose. It might be a question buried in a list of activities for which a respondent is asked to indicate whether she participates in them “frequently,” “sometimes,” “rarely,” or “never.” The directed-answer item in the list is worded, for example: “*This is a control question. Please mark “Rarely” and move on.*” Respondents who click on any option other than “rarely” are deemed to be inattentive to that question. (How to deal with evidence of inattentiveness is discussed in a later section.)

The three types of questions just given are exemplars of tools for detecting careless responding. There is no established theoretical justification for any particular number of such questions to include, or where to place them. A decision on the “right” number should take into account the length of the survey itself. There is no suggestion that all should be employed. Due diligence is satisfied by any quality control measure that detects and controls for inattention.

C. Addressing Cheating

Cheating involves attempting to improve answers by consulting other sources, or going back to change earlier answers to ones that do not reflect true “first impression” (as trademark surveys generally want to capture).

Addressing the risk of respondents’ returning to earlier questions is the easiest to address. The survey simply needs to be programmed accordingly, to disallow returns to earlier screens.

Addressing the risk of respondents’ consulting extraneous sources can be addressed in part by putting a reasonable time limit on the survey. A reasonable time limit would not capture instances of all “consulting activity” that the respondent may engage in, but it will contribute to quality control. The time limit can be pre-tested in advance to ensure that it is indeed “reasonable” for a wide range of reading and typing speeds.⁷² The authors simply ask participants not to consult other sources, and require them to confirm that they will not, as part of the introduction to the survey. Wording can be used such as this: “*Do you agree to answer each question truthfully and thoughtfully, and to not consult any other source of information, when completing this survey?*”

71. See Michael R. Maniaci & Ronald D. Rogge, *Caring about Carelessness: Participant Inattention and its Effects on Research*, *J. Research Personality* 48, 61-83 (2014).

72. Pre-testing is a standard step for any survey, and is typically done on a small sample that is nonetheless large enough to make statistical projections about the magnitude of any problem detected with the pretest.

It is important to note that inclusion of questions attempting to address faking, inattention, and cheating does not appear to interfere with legitimate survey results.⁷³

VII. POST SURVEY VERIFICATION AND ANALYSIS

The Post-Survey stage permits a final analysis of the integrity of the data, and an opportunity to remove suspicious response records.

Inattentive responders can be eliminated at this stage, if they were not eliminated earlier, based, say, on their “Attentiveness Score” described in the previous section. If such a procedure is applied, the number of instances of inattention for purposes of disqualification should be decided in advance, and the “rule” strictly applied. The expert should then also verify that the sample profile has not been distorted in the process. For example, if all eliminated people are under 25 years of age, then one might suspect that the elimination procedure itself has some built-in bias that needs to be rectified. It has indeed been found that there is a relationship between inattention and certain demographics.⁷⁴ To assure the integrity of the elimination process, one could present to the court a comparison of responses of the retained survey sample with those of the eliminated respondents.⁷⁵ If there are no significant differences, then nothing has been lost in the elimination process. Evidence of differences between the response records eliminated and those retained would confirm to the court the wisdom of the expert’s quality control procedures. If an opposing expert is not similarly diligent, and presents opposite survey results, the court may prefer the survey governed by superior quality controls.

A final verification procedure is to conduct telephone confirmation of 10% to 15% of respondents, by calling them directly. Access to this verification procedure requires the step described earlier, of allowing participation only by people willing to give their phone numbers for a possible future verification call. The interviewer making the calls would confirm that it was indeed the panel member who personally answered the survey, and that the respondent indeed met the survey qualifications. Verifying between

73. Adam J. Berinsky, Michele F. Margolis & Michael W. Sances, *Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys*, 58 Am. J. Pol. Sci. 744, 739 (2013) (hereinafter, “Berinsky et al.”).

74. Older respondents (compared to younger) and women (compared to men) are significantly more likely to pass attention screeners. In some instances, passage rates for attention screeners vary based on the respondent’s race. See Berinsky et al., p. 748, *supra* note 73.

75. *Id.*

10% and 15% of respondent information in this manner is consistent with verification levels required of telephone surveys in the past.⁷⁶

VIII. CONCLUSIONS

Internet surveys are increasingly used by experts in trademark disputes. While such surveys now enjoy cautious acceptance by the courts, they are vulnerable to criticism—as surveys have always been—by avid opposing experts. Features of Internet surveys particularly at risk for criticism include the inability to obtain a perfect random sample, and the power of participants to fake their responses, speed through surveys without attention, and “cheat” on their answers.

Internet survey reliability is still evolving. While Internet survey panels of volunteers are not a random sample of the population, investment in increasingly sophisticated, multiple recruiting methods⁷⁷ holds out hope for continuous improvement in survey reliability. In any event, Internet surveys can readily capture a demographically representative sample by design, obtain better response rates than telephone surveys, and reach across a much broader geography than mall samples.

Interventions under the control of the expert to enhance both reliability and validity can be implemented in three stages:

1. Choosing a panel company that is likely to deliver quality information about panel participants, and that employs sound sampling procedures;
2. Including questions within the body of the survey to check for faking, inattention, and cheating; and
3. Engaging in post-survey validation and analysis to strengthen the integrity of the data and provide assurance of legitimate respondent identities.

If best practices are in place in these areas, then trust by courts in the integrity of Internet surveys is in order. *Doveryai, no Proveryai*. Trust but verify.

76. See, e.g., Ruth M. Corbin & Fiona Isaacson, *Surveys on a Tightrope—the Convergent Validity Net*, 24 *Intell. Prop. J.*, 265-78 (2012).

77. One of the earliest steps to progress, described by panel company Probit at <https://www.probit.ca/what-we-do/probability-based-panels/>, is recruiting panel members by random-digit-dialing of phone numbers, which captures a random sample of listed, unlisted and cell-phone numbers. Multiple-recruiting methods for health-related studies are reviewed in T.S. Lane, J. Armin & J. S. Gordon, *Online Recruitment Methods for Web-based and Mobile Health Studies: A Review of the Literature*, *J. Med. Internet Res.* 2015 17(7).