

ESTTA Tracking number: **ESTTA702251**

Filing date: **10/14/2015**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE TRADEMARK TRIAL AND APPEAL BOARD

Proceeding	91200832
Party	Plaintiff Briggs & Stratton Corporation and Kohler Co.
Correspondence Address	ROBERT N PHILLIPS REED SMITH LLP 101 SECOND ST, STE 1800 SAN FRANCISCO, CA 94105 UNITED STATES ipdocket-chi@reedsmith.com, nborders@reedsmith.com, robphil- lips@reedsmith.com, knowakowski@whdlaw.com, dkalahele@rddsmith.com, ebridge@whdlaw.com, mgi
Submission	Plaintiff's Notice of Reliance
Filer's Name	Robert N. Phillips
Filer's e-mail	robphillips@reedsmith.com, sherring@reedsmith.com, dkala- hele@reedsmith.com
Signature	/s/ Robert N. Phillips
Date	10/14/2015
Attachments	Opposers' Eleventh Notice of Reliance.pdf(1133398 bytes)

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE TRADEMARK TRIAL AND APPEAL BOARD**

BRIGGS & STRATTON CORPORATION)	
)	
Opposer,)	Opposition No. 91200832 (Parent)
vs.)	
)	
HONDA GIKEN KOGYO KABUSHIKI KAISHA,)	
)	
Applicant.)	
)	
KOHLER CO.)	
)	
Opposer,)	Opposition No. 91200146
vs.)	
)	
HONDA GIKEN KOGYO KABUSHIKI KAISHA,)	
)	
Applicant.)	
)	

United States Patent and Trademark Office
Trademark Trial and Appeal Board
P.O. Box 1451
Alexandria, Virginia 22313-1451

OPPOSERS' ELEVENTH NOTICE OF RELIANCE

Pursuant to 37 C.F.R. § 2.122 and TBMP § 704.08(a), Opposers Briggs & Stratton Corporation (“Briggs”) and Kohler Co. (“Kohler”) (collectively, “Opposers”), by and through their attorneys, hereby submit their Eleventh Notice of Reliance. Opposers’ Eleventh Notice of Reliance contains the following attached materials:

1. Diamond, Shari Seidman. “Control Foundations: Rationales and Approaches.” *Trademark and Deceptive Advertising Surveys, Law, Science, and Design* (2012): pp. 201-216 included as **Exhibit RR**.

This printed publication is relevant to the issue of lack of secondary meaning of the Applicant's trademark. Specifically, this printed publication is authored by a leading authority on survey methodology, and addresses the proper selection and use of a control in a trademark survey.

The attached exhibit is marked with sequential page numbers in the upper right corner in the form of [Exhibit Letter – Page Number]. For example, the first page of Exhibit RR is “RR-1” and so forth. Where the upper right corner of a document does not allow for such marking, page numbers will be located along the right hand margin of the document. For briefing purposes, material within this Notice of Reliance will be identified as “O11NOR” (an acronym for Opposers' Eleventh Notice of Reliance). For example, a reference to the first page of Exhibit RR in the trial brief would be O11NOR RR-1.

Respectfully Submitted,

Dated: October 14, 2015

By: /s/ Robert N. Phillips
Robert N. Phillips
Seth B. Herring
Reed Smith LLP
101 Second Street
San Francisco, CA 94105

Attorneys for Opposer Briggs & Stratton
Corporation

Dated: October 14, 2015

By: /s/ Kenneth Nowakowski
Kenneth Nowakowski
Elizabeth Townsend Bridge
Melinda Giftos
Whyte Hirschboeck Dudek S.C.
535 East Wells Street, Suite 1900
Milwaukee, WI 53202

Attorneys for Opposer Kohler Co.

CERTIFICATE OF SERVICE

I hereby certify that a true copy of the foregoing OPPOSERS' ELEVENTH NOTICE OF RELIANCE was served via first class mail, postage prepaid, this 14th day of October, 2015 upon:

Vinita Ferrera, Esq.
John Regan, Esq.
Carrie Seares, Esq.
Sarah R. Frazier, Esq.
Shira C. Hoffman, Esq.
Wilmer Cutler Pickering Hale and Dorr LLP
60 State Street
Boston, Massachusetts 02109-1800
Phone: (617) 526-6448
Fax: (617) 526-5000

Kenneth R. Nowakowski
Elizabeth Townsend Bridge
Melinda Giftos
Whyte Hirschboeck Dudek S.C.
535 East Wells Street, Suite 1900
Milwaukee, WI 53202
Phone: (414) 273-2100
Fax: (414) 223-5000

/s/ Deborah L. Kalahale

Deborah L. Kalahale

EXHIBIT RR

Trademark

and

**Deceptive
Advertising
SURVEYS**

Law, Science, and Design

EDITED BY

Shari Seidman Diamond
and Jerre B. Swann



Section of
Intellectual Property Law

CONTROL FOUNDATIONS: RATIONALES AND APPROACHES

9

By Shari Seidman Diamond

INTRODUCTION

The Lanham Act explicitly specifies that a trademark owner who is claiming infringement must prove that consumers are likely to be influenced by the actions of the alleged infringer. That is, the trademark owner must show that the infringer's use "is likely to cause confusion, or to cause mistake, or to deceive. . . ."¹ Similarly, the statute requires the plaintiff alleging deceptive advertising to show misrepresentation likely to damage the plaintiff.² Analogous language appears in the statute concerning dilution: the plaintiff must show that the use "is likely to cause dilution. . . ."³ Thus, causation is an unambiguous requirement in each instance. It is perhaps surprising, therefore, that early trademark surveys did not recognize and address this statutory requirement.

1. 15 USC § 1125(a)(1)(A)(2011).

2. 15 USC § 1125(a)(1)(B)(2011). Assessing consumer reaction is necessary in determining whether misrepresentation has occurred, that is, that the consumer is likely to be deceived by the advertising message, unless the message is literally false or necessarily false by implication. *B. Sanfield, Inc. v. Finlay Fine Jewelry Corp.*, 168 F.3d 967, 971 (7th Cir. 1999).

3. 15 USC § 1125(e)(1)(2011).

The author is grateful to Raquel C. Rodriguez at Brinks Hofer Gilson & Lione for the research assistance she provided on this project when she was a law student at Northwestern University.

It is not as if the methodology suitable for conducting survey-experiments that could assess causation was unknown, or even obscure or impossible to implement. The underpinnings of the method go back at least to the publication of *The Design of Experiments* in 1935.⁴ Methods of experimental design have been the standard fare in both laboratory and field settings for many years, used in areas as diverse as medical research (does a drug or other medical treatment produce improvements in health?) and agricultural studies (is a fertilizer effective in promoting plant growth?). Yet parties and experts submitting surveys as evidence in trademark and deceptive advertising cases, perhaps failing to recognize that the relevant survey question was a causal one, regularly failed to include the controls that are crucial elements in experimental designs.⁵ Courts in turn admitted and relied on results from those control-absent surveys. As trademark and deceptive advertising surveys have evolved in the past 20 years, and particularly in the wake of growing federal court sophistication about surveys, the tide has turned.⁶ Although a few courts in recent years have been willing to admit and rely on surveys that lacked controls,⁷ the absence of a control is more often recognized as a fatal weakness.⁸

A similar need for a standard for comparison in the form of a control cell or cells or in the form of a control stimulus or stimuli arises in surveys assessing genericism or secondary meaning. We need to assess whether consumer responses in the survey reflect the construct of interest (e.g., for secondary meaning, that the consumer believes that the mark indicates that the product comes from a single source; for genericism, that the consumer believes that the mark as applied to the product indicates that it is a brand rather than a "common" name) or whether they have been produced by noise due to guessing or some other source (e.g., a belief by some consumers that all vodka or tomato sauce is made or authorized by the same company).

The second section of this chapter begins by reviewing in detail the threats to valid inference posed when a control is absent. The third section then shows how including an appropriate control cell can eliminate nearly all competing explanations

4. The author was British statistician Ronald Fisher, called "the father of modern statistics and experimental design" by Richard Dawkins at www.edge.org (2010).

5. For an early discussion of this neglect in cases of deceptive advertising, see Shari Seidman Diamond, *Using Psychology to Control Law: From Deceptive Advertising to Criminal Sentencing*, 13 *LAW & HUM. BEHAV.* 239-52 (1989).

6. Shari Seidman Diamond, *Reference Guide on Survey Research 400*, in *REFERENCE MANUAL ON SCIENTIFIC EVIDENCE* (3d ed., Fed. Jud. Ctr. 2011) (documenting growth of surveys with control groups in Lanham Act cases).

7. See, e.g., *Ironclad v. Poly-America*, 2000 U.S. Dist. LEXIS 10728 *23-25 (N.D. Tex. 2000) (although the court viewed the lack of a control as "the most significant challenge to the survey evidence" and discounted its weight, the court "yet still gives [the survey] results some consideration").

8. See, e.g., *Procter & Gamble Pharm., Inc. v. Hoffman-LaRoche, Inc.* 2206 U.S. Dist. LEXIS 64363 *25 (S.D.N.Y. 2006); *Simon Property Group L.P. v. mySimon, Inc.*, 104 F. Supp. 2d 1003, 1045-51 (S.D. Ind. 2000); *National Football League Properties, Inc. v. ProStyle, Inc.*, 57 F. Supp. 2d 665, 668 (E.D. Wis. 1999); *P&G v. Ultreo, Inc.*, 574 F. Supp. 2d 339, 351-52 (S.D.N.Y. 2008); *Bracco Diagnostics, Inc. v. Amersham Health, Inc.*, 627 F. Supp. 2d 384, 448 (D. N.J. 2009).

for the results from the test cell by providing a standard of comparison against which to gauge the interpretation of the answers that the consumers give in the test cell. This chapter uses the standard terminology from the research design literature to unpack the important link between random assignment control methodology and *internal validity*⁹ (that is, to what extent is the measured response (e.g., level of consumer confusion) caused by exposure to the treatment (e.g., the allegedly infringing mark or the allegedly false advertising message?)), with examples from trademarks and deceptive advertising. It also considers how control questions as well as control cells can in some cases assist in ruling out competing explanations. The fourth section turns to the issues that arise in choosing an appropriate control. Although these choices tend to be highly case-specific, some general patterns can be described. Note that the mere presence of a control cell does not guarantee an adequate test of the relevant causal proposition: the stimulus selected for the control cell must be appropriate, as described below. Finally, the fifth section of this chapter discusses situations in which multiple controls may be required.

THREATS TO INTERNAL VALIDITY AND SOURCES OF MEASUREMENT ERROR

To answer the question, Does the allegedly infringing mark mislead the consumer as to source?, it is natural to expose the consumer to the mark and then ask the consumer questions about the mark. If that is all we do, whatever response the consumer gives to our questions could be caused by the mark itself, but it could also be due to a variety of other sources. Courts have recognized many of these explanations that threaten the validity of the survey under the general category of noise,¹⁰ an appropriate if general description. The "noise" interferes with the signal (e.g., actual percentage of consumers who are misled) that the consumer's answer to the question would provide in the absence of noise. Two general categories of noise can cause interference. The first is noise that *systematically* distorts the pattern of responses, as, for example, when consumers have preexisting views about the product and respond based on those views rather than responding to the content of the commercial being tested. The second category of noise is noise that adds a *random* component to a respondent's answers, as when the respondent who is presented with a choice of possible answers chooses one of them randomly—that is, offers a pure guess. Whether noise is the

9. DONALD T. CAMPBELL & JULIAN C. STANLEY, EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR RESEARCH (1966).

10. See, e.g., *Novartis Consumer Health, Inc. v. Johnson & Johnson-Merck Consumer Pharm. Co.*, 290 F.3d 578 (3d Cir. 2002); *Winning Ways, Inc. v. Holloway Sportswear, Inc.*, 913 F. Supp. 1454, 1475 (D. Kan. 1996).

result of systematic or random distortion,¹¹ it can compromise the results of a survey. Following are the primary forms of noise that impair inference in the absence of controls. It is important to recognize that although these forms of noise can seriously undermine the value of a survey, appropriate controls offer an effective way to remove most of these threats.

Preexisting Beliefs

All respondents answer questions against a background of preconceived notions. These may include general beliefs about the nature of trademarks (e.g., all products in a category are really made by the same company) and commercials (e.g., all commercials claim that their products have positive features). They may include specific beliefs about the particular characteristics of a mark or a product, such as an awareness of the name of the best-selling brand in a particular product category or a belief that a particular substance is harmful. These preconceived notions naturally condition how the respondents will interpret and answer questions about the trademark or commercial they are shown in the course of an interview. Some surveys attempt to reduce the impact of preexisting impressions on respondents' answers by instructing respondents to focus solely on the stimulus as a basis for their answers. Thus, the survey includes a preface (e.g., "based on the commercial you just saw") or directs the respondent's attention to the mark at issue (e.g., "these stripes on the package"). Such efforts are likely to be only partially successful. It is often difficult for respondents to identify accurately the source of their impressions.¹² The more routine the idea being examined in the survey, the more likely it is that the respondent's answer is influenced by preexisting impressions, rather than by the actual content of the commercial message or trademark being evaluated. In order to know whether an allegedly infringing trademark or an allegedly deceptive commercial is the cause of the respondent's answer, it is necessary to rule out these preexisting notions as alternative explanations for the answer. That is, we need a control that will provide a measure of preexisting notions.

Yea-Saying

Acquiescence bias occurs because subsets of individuals, often called yea-sayers, have a "tendency to endorse any assertion made in a question, regardless of its content."¹³ This occasion for biased endorsement arises whenever a question is framed as a request for agreement or disagreement, as when the response choices are agree/disagree, true/false, or yes/no.¹⁴ For example, questions like, "Have you ever heard

11. Statisticians divide measurement error into two types: bias, which takes the form of systematic distortion, and sampling error, which results in random distortion; See David H. Kaye & David A. Freedman, *Reference Guide on Statistics* at 283, 296, in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE (3d ed., Fed. Jud. Ctr. 2011).

12. See Richard E. Nisbett & Timothy D. Wilson, *Telling More Than We Can Know: Verbal Reports on Mental Processes*, 84 *PSYCHOL. REV.* 231 (1977).

13. Jon A. Krosnick, *Survey Research*, 50 *ANN. REV. PSYCHOL.* 537, 552 (1999).

14. Diamond, *supra* note 6 at 394.

survey.
ence of
riously
remove

i. These
cts in a
mercials
beliefs
ness of
f that a
on how
mercial
impact
o focus
preface
attention
likely to
curately
l in the
existing
demark
k or an
cessary
That is,

-sayers,
; of its
framed
; agree/
r heard

/stematic
David A.
VIDENCE

! Reports

of Mercury as a brand of automobile?" may attract agreement even from respondents who are actually unaware that Mercury is a brand of automobile. As courts have recognized, the presence of yea-sayers artificially inflates the rate of agreement.¹⁵ Moreover, the effect is not insignificant: the bias has produced an inflation effect of 10 percent across a number of studies.¹⁶

Guessing That Produces Random Error

Respondents in trademark surveys are typically asked to avoid guessing, but the survey setting itself invites respondents to provide answers to the questions they are being asked and, having agreed to participate in the survey, many respondents will make an effort to offer answers. Further, respondents want to appear knowledgeable, and thus "don't know" is generally a less attractive option if other choices are readily available. Either guessing or choosing "don't know" is likely to be stimulated if the question or the response options are challenging to understand, as occasionally occurs when the survey is testing the message that consumers received from an allegedly deceptive commercial. The respondent may guess or choose "don't know" rather than engage in carefully considering each of the alternative response choices about the message being conveyed by the commercial. It is difficult to know how much guessing is occurring in a particular survey without some additional evidence. One reason why surveys often include multiple questions aimed at measuring the same attitude or belief is that consistent responses by individuals across several questions on the same topic provide a more stable measure, as well as evidence that the answers are not simply a result of guessing.

Many of the questions asked in trademark surveys present respondents with two choices: Do you think X and Y are put out by the same or by different companies? Do you think that Z is put out by one company or by more than one company? If respondents were merely to guess, each of the two choices would attract half of the responses. As a result, even if a "don't know" option is added to reduce guessing, the pattern of responses will be hard to interpret. That is, how many of those responding that Z is put out by one company were responding to the mark and how many were merely guessing? Without a control for guessing, we cannot tell.

Guessing That Produces Systematic Error

This form of guessing can appropriately be characterized as "guided guessing." It occurs if questions or response choices contain cues that suggest a particular response. The extreme form of suggestion is the familiar leading question that makes it clear that a particular response is expected or preferred. For example, respondents in a survey were asked, "To the best of your knowledge, was the Donkey Kong game made with the approval or under the authority of the people who produce the King Kong movies?"¹⁷

15. *Pharmacia Corp. v. GlaxoSmithKline Consumer Healthcare, L.P.*, 292 F. Supp. 2d 594, 601 (D. N.J. 2003).

16. Krosnick, *supra* note 13.

17. *Universal City Studios, Inc. v. Nintendo Co.*, 746 F.2d 112, 118 (2d Cir. 1984).

The court rejected the question as leading, observing that “[A] survey question which begs its answer cannot be a true indicator of the likelihood of consumer confusion.”¹⁸ Questions can also provide more subtle cues as to “correct” or preferred answers. By providing a frame for responses, a question can direct the respondent to answer with the categories mentioned in that frame. For example, when Walmart claimed that Charles Smith was infringing and diluting its trademarks through Smith’s sale of T-shirts that Smith sold only on the web, Walmart’s survey asked respondents “What company or store puts out this product?”¹⁹ As the court recognized, “the disputed questions improperly led respondents to limit their answers to companies or stores.”²⁰

Finally, earlier questions in a survey can provide a context for a question that influences the pattern of answers that respondents are likely to give to the questions that follow. For example, in *P&G v. Ultreo*, the plaintiff sued a competitor for allegedly deceptive advertising about the use of ultrasound technology in its new power toothbrush.²¹ Ultreo had advertised its new power toothbrush in part based on the ultrasound feature of the toothbrush and the plaintiff charged that the implied claims about the effects of the ultrasound feature were deceptive. The court pointed out that the plaintiff’s survey questions asking respondents “What does the ultrasound do?” and “What is the benefit of ultrasound?” focused the attention of the respondents on ultrasound.²² The court concluded that these questions may have stimulated ultrasound responses that did not reflect the message respondents took from the commercial.²³

A CLOSER LOOK AT SURVEY DESIGN WITHOUT AND WITH CONTROLS

Without Controls

Suppose that we show the consumer a can that is the size and shape of most cans that contain carbonated beverages. The can is red with white block letters saying cola. We ask consumers, Who do you believe puts out this product? (the traditional Eveready approach).²⁴ Some percentage of respondents may assume that Coca-Cola is the source (or that the producer had to get permission from the makers of Coca-Cola). Although we may caution respondents not to guess, as well as give them the option of saying “I don’t know,” respondents in a survey want to be helpful and to appear knowledgeable. If 28 percent of respondents in a survey say Coke or Coca-Cola, should this result be viewed as sufficient evidence to support a claim of likely

18. *Id. See also*, 1-800 Contacts, Inc. v. Lens.com, Inc., 2010 U.S. Dist. LEXIS 132948 at *22–23 (D. Utah 2010) (“the form of the question strongly suggested the response.”).

19. *Smith v. Wal-Mart Stores, Inc.*, 537 F. Supp. 2d 1302, 1319 (N.D. Ga. 2008).

20. *Id.* at 1332.

21. *P&G v. Ultreo, Inc.*, 574 F. Supp. 2d 339 (S.D.N.Y. 2008).

22. *Id.* at 352.

23. *Id.*

24. *Union Carbide Corp. v. Eveready, Inc.*, 531 F.2d 366 (7th Cir. 1976).

confusion? The problem is that a variety of legally irrelevant reasons may account for the Coke responses. For example, the untrademarked color red may have led some respondents to say Coke. Others may have chosen Coke because, as the product leader in the cola category, Coke was the brand name most available to respondents and thus the most likely brand to guess.

The evidence would be even further undermined if the question instead was: which of the following, if any, puts out this product: Coke, Pepsi, 7Up, other? If a respondent had absolutely no idea and simply guessed, 25 percent of respondents would say Coke. Providing an explicit "don't know" option typically reduces guessing by 20 to 25 percent without substantially changing the distribution of the respondents selecting each of the listed choices.²⁵ Subtracting the expected responses that would be diverted to the "no opinion" option, we would expect to obtain 25 percent minus 6.25 percent (25% of 25% = 6.25%), that is, 18.75 percent Coke responses *if respondents were merely guessing*.²⁶ If we add to that the respondents who were more likely to choose Coke merely because of its large market share—that is, we take into account systematic as well as random guessing—the percentage selecting Coke would go higher even if no confusion due to trademark infringement was actually occurring. Finally, suppose the question was: Is the product you just saw put out by the Coca-Cola Company or by some other company? Reducing the number of choices to two (Coke or other) would inflate the base rate for guessing even further. Thus, the survey evidence of 28 percent Coke responses that, at first blush, might seem to reveal substantial evidence of likelihood of confusion, dissolves on closer examination.

This survey design—without a control—is what is referred to as a One-Group Posttest-Only Design.²⁷ Diagrammatically, it can be represented as follows:

X O

This diagram depicts time moving from left to right; X = the test stimulus or treatment (in this case the red cola can with white block lettering); and O = the observation (in this case the questions asked in the survey about the source of the cola).²⁸

To see how the weaknesses of this design also apply in a deceptive advertising context, consider X as an allegedly deceptive commercial claiming the virtues of an antacid product and allegedly implying (although not explicitly stating) that the ingredient beryllium used in the product of its chief competitor is harmful. In fact, the competitor's product does contain beryllium, but there is no scientific evidence that beryllium in the quantities involved is in fact harmful. Consumers are shown the

25. HOWARD SCHUMAN & STANLEY PRESSER, *QUESTIONS AND ANSWERS IN ATTITUDE SURVEYS: EXPERIMENTS ON QUESTION FORM, WORDING AND CONTEXT* 113-46 (1981).

26. The court in an early case acknowledged the problem of accounting for guessing that would produce 20% of respondents endorsing each of five possible choices. *American Home Products Corp. v. Johnson & Johnson*, 654 F. Supp. 568 (S.D.N.Y. 1987).

27. THOMAS D. COOK & DONALD T. CAMPBELL, *QUASI-EXPERIMENTS: NONEQUIVALENT CONTROL GROUP DESIGNS* (1979) at 96.

28. *Id.* at 95-96. The notation was originally developed by D.T. Campbell and J.C. Stanley in their classic monograph *EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR RESEARCH* (1966).

commercial and then—the observation O in the diagram above—the follow-up survey questions them on their beliefs about the competitor's product and its ingredients, based on the commercial. If 28 percent of respondents report that the beryllium in the competitor's product makes it dangerous, has the survey demonstrated, as the plaintiff is required to prove in a case of deceptive advertising based on implied falsity, that the commercial has caused this belief? The problems we saw in the earlier example infect such an inference here as well.

Respondents come to products, trademarks, trade dress, and commercials with expectations and beliefs and those expectations and beliefs influence their survey responses. Consumers who have previously been exposed to negative news stories on the dangers of beryllium and those who assume it is dangerous based only on the name "beryllium" (e.g., sounds like radium) view the commercial with a jaundiced eye. Even if the commercial merely listed the ingredients of the competitor's product with no further suggestion about beryllium dangers, on being questioned after viewing the commercial those respondents may report negative feelings about the competitor's product due to its beryllium content. Neither the researcher using a research design that lacks a control nor the court evaluating the results can know how many of the responses reporting that beryllium in the competitor's product makes it dangerous are due to these preexisting beliefs, how many are due to spontaneous responses to the ingredient's name, and how many are due to a false implied message of dangerousness in the commercial. Note that although the survey question asks the respondents to answer "based on the commercial," it is difficult for respondents to follow this direction because it requires respondents to be able to accurately gauge the basis of beliefs that may have multiple sources, a standard problem of human inference.²⁹ Even highly motivated respondents may be unaware of what influences their response.³⁰

Now suppose that we change the design. If preexisting attitudes and beliefs are a problem, can we simply measure those and then "net out" those responses in evaluating the post-exposure survey results? The design, labeled a One-Group Pretest-Posttest Design,³¹ becomes:

$$O_1 \quad X \quad O_2$$

That is, we measure at O_1 , before the commercial is shown to respondents, and at O_2 , after the commercial is shown. Can we then use O_2 minus O_1 to estimate the percentage of respondents deceived by the commercial, net of preexisting attitudes and beliefs? We have improved the ability to make valid inferences in one way by adding this pretest, but in doing so, we have introduced other threats to validity. The questions we have asked at O_1 have alerted respondents to the issues of interest (e.g., response to beryllium), potentially changing the way the respondents viewed the commercial, now sensitized to that issue. This priming that can affect how the respondent interacts with the commercial

29. Nisbett & Wilson, *supra* note 12.

30. *Id.*

31. Cook & Campbell, *supra* note 27 at 99.

represents priming that can affect precisely the kind of distortion that courts often appropriately object to when they complain that a survey did not appropriately replicate what would occur in the marketplace. Of course, no survey completely replicates market conditions, but one that introduces a change like the priming test in the One-Group Pretest-Posttest design leaves a trail of threats to valid inference in its wake. The result is that no reliable causal inference about the impact of X can be obtained.

With Controls

Fortunately, surveys in trademark and deceptive advertising litigation never need to rely on the weak preexperimental designs described above in order to test the relevant causal questions that arise in these cases. The primary solution is an experimental design in which survey respondents are randomly assigned either to the test cell or to a control cell (or one of a number of control cells).³² The genius of Ronald Fisher³³ was to recognize that equivalent groups of plots of land (or sets of people) could be created by randomly assigning them to conditions, thereby producing two (or more) sets of plots or people that did not initially differ. By subjecting one group to a treatment (the test cell) and the other(s) to no treatment or to different treatments (the control cell(s)), the only explanation for any difference observed between the groups on the post-test measure would be the treatment. The design rules out the possibility that, for example, preexisting attitudes or beliefs or differences in the composition of various plots of land could be responsible, because the groups on average (before treatment) were equivalent as a result of the random assignment to the test and control conditions. Using the same notation we used earlier, the basic randomized experimental research design becomes:

$$\begin{array}{l} R X O \\ R O \end{array}$$

where the R s indicate that units (people) are randomly assigned to either the test group (line 1) or the control group (line 2). Additional rows can be added to the design if additional control cells are added (e.g., one control group of respondents watches a commercial with the allegedly deceptive material removed, while a second control group of respondents watches no commercial at all). With this design, any differences in O become unambiguously attributable to the impact of the treatment. As we shall see below, precisely what it is about X that had the observed effect on the post-treatment measure can be a different question, but random assignment and the use of

32. Random assignment should not be confused with random selection. When respondents are assigned randomly to different treatment groups (e.g., respondents in one group watch an allegedly deceptive commercial while those in a second group watch no commercial or an acknowledged nondeceptive commercial), the procedure ensures that within the limits of sampling error the two groups of respondents will be equivalent except for the different treatment they receive (e.g., the different commercials they view). Respondents selected in a mall intercept study, as well as those selected in a probability sample, may be assigned randomly to different treatment groups. Random selection, in contrast, describes the method of selecting a sample of respondents in a probability sample. See Section III.C. in Diamond, *supra* note 6.

33. See text *supra* at note 4.

a control enable us to conclude unambiguously³⁴ that there was something about X that caused the post-treatment difference between groups.

CHOOSING APPROPRIATE CONTROLS

The general principle for choosing an appropriate control is easily stated: It should share as many characteristics with the experimental stimulus as possible, with the key exception of the characteristic whose influence is being assessed. The reason for this approach is that it peels away any confounding influences that threaten the inference that the experimental stimulus caused the differences between the test and control groups on the post-treatment measure. The devil is of course in the details. How much sharing is enough? On what dimensions? How much sharing is too much? Judge Posner once opined that survey research involves "black arts," suggesting that devious manipulation might explain the differences between survey designs when opposing experts present surveys that produce different results, each favoring the particular expert's client.³⁵ While survey designs are not always optimal and experts are not always pure of heart, an alternative explanation to pure partisan manipulation in many cases is that good survey design requires the employment of "analytic" rather than "black" arts. The key is to analyze what needs to be present in the control and what needs to be absent, and to identify a control stimulus or set of stimuli that meet both requirements.

We begin with a recent case involving a relatively simple example of a control group survey design. RE/MAX sued TREND SETTER realty for trademark infringement, alleging that the red, white, and blue yard sign used by TREND SETTER was likely to cause confusion with RE/MAX's own red, white, and blue sign.³⁶ The RE/MAX sign consisted of three horizontal bars with red at the top, white in the middle, and blue on the bottom of the sign. The TREND SETTER sign used a similar color scheme and horizontal approach to the use of the three colors, but the white bar protruded into the red bar at the top to form a roof-sloped image.

Survey respondents viewed either the TREND SETTER sign or a control sign. The control sign was identical to the test sign except that the red and blue horizontal bars were modified so that they were white rather than red or blue.³⁷ The white roof silhouette was preserved with a blue line outlining the roof. Over one-fourth (25.3 percent) of the individuals who viewed the test sign said either that RE/MAX was the company being promoted or advertised by the sign, or that the company being promoted or advertised by the Trend Setter Realty sign was affiliated or connected

34. Within the limits of sampling error. See Kaye & Freedman, *supra* note 11 at 296.

35. Indianapolis Colts v. Metro Baltimore Football Club Ltd. P'ship., 34 F.3d 410, 416 (7th Cir. 1994).

36. RE/MAX Int'l, Inc. v. Trendsetter Realty, LLC, 655 F. Supp. 2d 679 (S.D. Tex. 2009).

37. The white print on the red and blue portions of the test sign was changed to red on the control sign so that it would be visible when the background of the red and blue portions became white (personal communication, Robert Peterson, April 28, 2012).

with RE/MAX or would have to get permission or approval from RE/MAX to use the sign. Only 2.7 percent of those who saw the control sign identified RE/MAX as the source. The net confusion rate was thus 25.3 percent minus 2.7 percent = 22.6 percent and the court found that the survey evidence was "sufficient to support the conclusion that there is actual confusion."³⁸

By removing the color of the red and blue bars from the color scheme of the Trend Setter sign and asking the same questions to respondents in the control cell, the control cell provided an estimate of the extent to which respondents to the test sign were answering with the name "RE/MAX" for reasons other than the RE/MAX trademark (e.g., merely because it was more familiar through its dominant standing in the real estate market). Although another version of the control sign might have been used (e.g., one with a different color scheme), the choice of the control image was reasonable and permitted an unambiguous inference that the difference in the color schemes used in a horizontal pattern on the two signs was responsible for the difference in the rate of RE/MAX responses.

The court was further impressed by the verbatim responses to the follow-up questions, which asked the respondents to explain why they gave their answers: "Why do you say that?" and "What other reasons?" Many of the respondents mentioned the colors on the sign. Several points are relevant here. First, even in this relatively straightforward case in which it was easy to indicate the characteristics of the mark being tested and there were few obvious competing characteristics on the stimulus that might have drawn the respondent's attention, responses to the "why?" questions tended to be incomplete. Here, 58 percent (22/38) of the respondents who said RE/MAX gave color as an explanation, while 42 percent (16/38) did not mention color in their responses.³⁹ Moreover, "reasons" can be misleading, that is, they may not be reasons at all. In the RE/MAX survey, some respondents gave "reasons" that did not relate to the RE/MAX logo (e.g., "because it had a house on the sign"⁴⁰—in fact, there is no house on the RE/MAX logo) or otherwise did not make sense (e.g., "I remember seeing an R on the top right corner"⁴¹—in fact, there was no R in the top right corner of the sign). The reason why such pseudo-reasons may occur is that respondents who are questioned about their reasons for an answer will search for a plausible explanation that may or may not be the reason for that earlier response. Although people are often able to *justify* their positions when asked to explain their actions, such post-hoc explanations can only imperfectly capture the reasoning that actually produced their answers.⁴² The best method for determining the cause of a confusion response is a survey design with a tight control that directly isolates the explanatory feature by showing that the answer changes when the feature is changed or is no

38. RE/MAX, 655 F. Supp. 2d at 706.

39. Computed from a review of verbatim responses in Expert's report, 2007 Misc. Filings 72426, 2009 Misc. Filings LEXIS 1632 at *13-18 (S.D. Tex. 2009).

40. *Id.* at *14.

41. *Id.* at *15.

42. Nisbett & Wilson, *supra* note 12.

ut X that

t should
with the
e reason
aten the
test and
: details.
o much?
ting that
as when
ring the
l experts
pulation
s" rather
ontrol and
hat meet

i control
ademark
TREND
und blue
p, white
n used a
, but the

rol sign.
orizontal
ite roof
th (25.3
AX was
y being
nnected

ir. 1994).

ontrol sign
(personal

longer present. That is precisely what the pattern of responses (that is, few RE/MAX responses) in the control cell supplied in the RE/MAX case. The net percentage of RE/MAX responses (test cell percentage minus the control cell percentage) supplied the most telling evidence of the level of confusion caused by the color pattern used by TREND SETTER on its yard sign. When the characteristics of a mark or the allegedly deceptive message in a commercial are more complex, the features of a good control that will accomplish this purpose become more complex as well. We next describe four features that characterize a good control stimulus:

1. A good control stimulus shares features with the allegedly infringing mark or deceptive advertisement, other than those alleged to be infringing or deceptive, that might affect responses;
2. it does not contain cues that will artificially depress confusion responses by leading the respondent in another direction;
3. if it is a control for an allegedly infringing mark, it should appear to be a plausible member of the same product category; and
4. if it is a control for an allegedly infringing mark, it should not itself be an infringing mark.

Selecting a Control That Shares Features with the Test Stimulus That Are Not at Issue

The control stimulus should retain as many noninfringing characteristics of the test stimulus as possible.⁴³ In the RE/MAX example described earlier, the control sign shared crucial features with the test sign. The control sign preserved the roof image of the TREND SETTER design that appeared in the test cell and that might have encouraged respondents to name a familiar real estate company. In addition, both the test and control signs included the TREND SETTER name, so that the presence of the TREND SETTER name could not explain different responses in the test and control cells. These consistencies between the test and control signs made it possible to isolate the effect of the colored bar scheme on responses.

When an advertisement contains an allegedly misleading claim about the virtues of its product as compared with the product of a competitor, consumers are likely to report that the commercial made a superiority claim, but that report may not be based on an implied message contained in the commercial about the actual characteristics or performance of the product relative to competing products. Instead, it may simply reflect a generalized response to comparative advertising: in a comparative advertisement, consumers expect advertisers to extol the virtues of their products relative to those of competitors. Thus, respondents in a survey who are questioned about an advertisement for a pain reliever may agree that the advertisement conveys a

⁴³ See, e.g., *Skechers USA, Inc. v. Vans, Inc.*, No. CV-07-01703, 2007 WL 4181677, at *8-9 (C.D. Cal. Nov. 20, 2007) (in trade dress infringement case, control stimulus should have retained design elements not at issue).

message that it reduces pain longer than does the pain reliever of its major competitor simply based on their expectations that a comparative commercial is likely to be making a superiority claim. Under these circumstances, the appropriate control commercial would be a comparative commercial that does not focus on superior duration of pain relief. Thus, in *Pharmacia v. GlaxoSmithKline*, the claim was that the manufacturer of a nicotine patch was falsely implying that the patch produced superior quitting efficacy.⁴⁴ The control commercial was not comparative. Indeed, while 92 percent of the respondents who saw the test commercial rated it as comparative, only 10 percent who saw the control commercial said it was comparative.⁴⁵ The court concluded that the noncomparative control ad failed to control for preexisting beliefs and expectations.⁴⁶ Respondents who were asked whether the commercial mentioned a difference in quitting efficacy could have been inclined to say that the comparative ad did make that claim, not as a result of the specific content of the comparative ad, but because they assumed that a comparative ad for the product would make such a claim. Note that an appropriate control commercial may not have been readily available. In that situation, it might have been possible to produce one by excising the allegedly misleading portion of the test commercial, or it may have been necessary to create a mock comparative commercial to act as the control.⁴⁷

Avoiding Cues in the Control That Artificially Depress "Confusion" Responses

It is often appropriate to use other existing brands in a survey that measures likelihood of confusion. The product or products must be chosen carefully to maximize similarity to the allegedly infringing product on all possible dimensions. The choice of a control brand should not artificially direct the respondent away from indicating that the control product is put out by the mark holder.⁴⁸ For example, the makers of NU FINISH car polish complained that FINISH 2001, a car polish put out by TURTLE WAX, infringed the NU FINISH trademark and trade dress.⁴⁹ The plaintiff's survey used an ARMOR ALL product as the control and found that few respondents said that it was put out by the company that made FINISH 2001; that is, there was little evidence of noise.⁵⁰ As the court recognized, the fact that ARMOR ALL, the maker's brand, was prominently displayed on the control product (unlike the way the small print name of the manufacturer appeared on the rear of the other bottles) made the

44. *Pharmacia Corp. v. GlaxoSmithKline Consumer Healthcare, L.P.*, 292 F. Supp. 2d 594, 597 (D.N.J. 2003).

45. *Id.* at 603, n. 4.

46. *Id.* at 604-05.

47. *Id.* at 604.

48. See e.g., *Louis Vuitton Malletier v. Dooney & Bourke, Inc.*, 525 F. Supp. 2d 576, 595 (S.D.N.Y. 2007) (underreporting of background "noise" likely occurred because handbag used as control was quite dissimilar in shape and pattern to both plaintiff and defendant's bags).

49. *Reed-Union Corp. v. Turtle Wax, Inc.*, 869 F. Supp. 1304 (N.D. Ill. 1994).

50. *Id.* at 1311.

ARMOR ALL product a choice that consumers would easily reject.⁵¹ As the court said, the “control product could not measure inherent confusion in the marketplace because it prominently identified the maker.”⁵² It is perhaps not surprising that an opposing survey using PRISM as the control product, a polish that did not display the maker’s name on its front label, produced substantially greater evidence of noise.⁵³

Selecting a Control That Is a Plausible Member of the Product Category

Control products may exist that have the needed characteristics for the survey, or they may be created to have them. In either case, they should be plausible members of the set of products at issue. For example, a survey was designed to test whether physicians were likely to believe that a new glaucoma medication TRAVATAN was produced by the company that manufactured the existing medication XALATAN.⁵⁴ The control, used by both the plaintiff’s expert in a dilution survey and the defense expert in a likelihood of confusion survey,⁵⁵ was LUMIGAN, another recent entry into the glaucoma medication market. All respondents in the likelihood of confusion survey were shown a page that resembled a magazine advertisement and contained the same illustration, but that varied the name of the product being advertised.

In contrast, a survey designed to test whether a new perfume bottle infringed the trade dress of Calvin Klein’s ETERNITY perfume used a control perfume bottle that failed to provide a plausible control.⁵⁶ According to the court, one reason for discounting the confusion survey was that it failed “to use a control that ‘could plausibly emanate from Calvin Klein.’”⁵⁷

Avoiding a Control That Is Itself Infringing

In an effort to control for all noninfringing elements of a mark, a survey may select a control that is itself infringing.⁵⁸ If so, the “control” provides no control at all because any “confusion” responses in the control may be the result of noise or actual confusion or a combination of both: there is no way to tell. Indeed, in a case involving a claim of trade dress infringement of Little Golden Books, a series of children’s storybooks, a control used in one of the surveys was, according to the court, “the one storybook on the market arguably more similar in trade dress to the Little Golden Book than [the defendant’s] product!”⁵⁹ As a result, the survey received no weight.

51. *Id.*

52. *Id.*

53. *Id.* at 1311–12.

54. *Pharmacia Corp. v. Alcon Labs., Inc.*, 201 F. Supp. 2d 335.

55. The author was an expert for the defendant in this case.

56. *Conopco, Inc. v. Cosmair, Inc.*, 49 F. Supp. 2d 242 (S.D.N.Y. 1999).

57. *Id.* at 254. *See also*, *THOIP v. Walt Disney Co.*, 788 F. Supp. 2d 168, 181 (S.D.N.Y. 2011).

58. *See e.g.*, the plaintiff’s critique of the defense survey in *Classic Foods Int’l Corp. v. Kettle Foods, Inc.*, 2006 U.S. Dist. LEXIS 97200 at *19 (C.D. Cal. Mar. 2, 2006) (what was characterized as “noise” in response to defense control could in fact be actual consumer confusion caused by the control stimulus).

59. *Western Publishing Co. v. Publications Int’l*, 1995 U.S. Dist. LEXIS 5917 at *29 (N.D. Ill. 1995).

Similarly, the defendant in a trade dress infringement case involving "Splenda" no-calorie sweetener produced a survey that purported to show an absence of confusion with the trade dress of the defendant, "Same," which was also packaged in a horizontal box with a yellow background and blue and white accents.⁶⁰ The defense survey used a "Sugar Twin" box as the control and found 22 percent minus 20.6 percent = 1.4 percent net confusion.⁶¹ The court rejected the evidence from the control on the ground that the control was itself infringing.⁶²

As these results reveal, before designing a survey and selecting a control, the expert should be as clear as possible on the elements that are included in the claimed trademarks and those that are unambiguously noninfringing. In some instances, it may be possible to identify the boundaries of the claim from the complaint. In others, it may be possible to fill in if there have been enforcement actions that have resulted in agreements as to acceptable modified uses of the mark (e.g., the "Fit 24 Club" and "The 24 Hour Gym" were recognized as acceptable uses of the trademark "24 Hour Fitness" for fitness facilities⁶³). To the extent that judges can assist in getting the parties to clarify just where the boundaries to the claimed trademark rights are located before the parties undertake any survey work, the surveys that are conducted will be better constructed and more likely to provide results that can withstand scrutiny and avoid ambiguity in the value of the information they can offer. Although it might be desirable to encourage even closer pre-survey narrowing of issues, such consultation is hard to produce. It did occur at least once in the case of *Sunamerica Corp. v. Sun Life Assur. Co. of Canada*.⁶⁴ The court in *Sunamerica* worked with the parties who then conducted a single survey. That level of collaboration has not, to this author's knowledge, been repeated.⁶⁵ The suggestion here is much narrower: encourage delineation of clear boundary information from the party claiming infringement, somewhat in the nature of an extended notice provision clarifying the scope of the claim.

THE ROLE OF MULTIPLE CONTROLS

To separate the effects of several factors or product features on likelihood of confusion, it may be advisable to use more than one control cell. This situation arose when the maker of M&M's chocolate candies alleged trade dress infringement based on the similar packaging of multicolored chocolate lentils used by the defendant on ROCKLETS candy packages. The plaintiff's survey included more than one control cell, including one in which the candy lentils seen through the package window

60. McNeil-PPC, Inc. v. Merisant Co., 2004 U.S. Dist. LEXIS 27733 at *11 (D. P.R. 2004).

61. *Id.* at *61.

62. *Id.* at *62. "Sugar Twin" had been marketed prior to the introduction of "Splenda" and thus was not subject to an infringement claim by the makers of "Splenda." *Id.* at *65.

63. 25 Hour Fitness USA, Inc. v. 24/7 Tribeca Fitness, LLC, 447 F. Supp. 2d 266, 280 (S.D.N.Y. 2006).

64. 890 F. Supp. 1559, 1570 (N.D. Ga. 1994).

65. Nor did agreement on the survey design prevent disagreement between the parties on the interpretation of the results. *Id.*

were multicolored, as they are on the M&M's packages, and a second in which the lentils were all white. Not surprisingly, fewer respondents who saw the second all-white lentils package answered that it was put out by the maker of M&Ms than did respondents in the cell in which the visible lentils were multicolored. The two controls provided the court with a range of estimates of noise, a baseline range against which to gauge the effect of the defendant's packaging on likelihood of confusion.

This chapter has focused primarily on control cell survey designs (that is, *between-group designs* that compare the responses of one set of survey participants with responses from another set). These are the strongest designs for causal inference because a respondent cannot be influenced by the content of material shown to respondents in other cells of the design. Of course, each additional cell in a between-group survey adds the expense of additional respondents. In contrast, minimal additional expense is entailed when the same respondents are asked to answer questions about multiple stimuli. These *within-group designs* are not suitable in some situations (e.g., in false advertising surveys), but are appropriate and standard fare in surveys assessing whether a mark is generic.⁶⁶ The other marks in the genericism survey provide a baseline of accepted brand and common names, so that the survey can reveal where the name at issue falls on the spectrum of tested names. A key threat to validity that arises in within-group survey designs is the danger of order effects.⁶⁷ It is therefore crucial that the order of presentation of stimuli in a within-group design be rotated.

CONCLUSION

Controls play a central role in enabling a survey to rule out threats to valid causal inference. A control group design that includes an appropriate control is the best way to ensure that noise from preexisting beliefs, yea-saying, and guessing (both random and biased) cannot explain away or undermine evidence of confusion or deception reflected in the responses of survey participants. An appropriate control or controls in a trademark survey, as in any experiment, should mimic the test stimulus as closely as possible without including features that are the focus of the infringement claim or are part of the message that is allegedly deceptive. A control should not be so different that it artificially lures respondents away from revealing the effect of noise. In an infringement case, the control should be a plausible member of the product class, but not so similar to the mark holder's product that it is itself infringing. Designing an appropriate control is thus an analytic rather than a mechanical task, one that is crucial in a valid trademark or deceptive advertising survey.

66. See Chapter 6, *infra*, by E. Deborah Jay

67. See HOWARD SCHUMAN & STANLEY PRESSER, QUESTIONS AND ANSWERS IN ATTITUDE SURVEYS: EXPERIMENTS ON QUESTION FORM, WORDING AND CONTEXT 23, 56-74 (1981); Jon A. Krosnick & Stanley Presser, *Questions and Questionnaire Design*, in HANDBOOK OF SURVEY RESEARCH 13 (James D. Wright & Peter V. Marsden eds., 2d ed. 2010) at 278-81.