

The Trademark Reporter[®]



The Law Journal of the International Trademark Association

A History of the Evolution of Likelihood of Confusion Methodologies
Jerre B. Swann

Why Ask Why? A Critical Assessment of an Historical Survey Artifact
R. Charles Henn Jr.

INTERNATIONAL TRADEMARK ASSOCIATION

675 Third Avenue, New York, NY 10017-5704

Telephone: +1 (212) 642-1700

email: wknnox@inta.org

OFFICERS OF THE ASSOCIATION

JOMARIE FREDERICKS*President*
DANA NORTHCOTT *President-Elect*
ELISABETH BRADLEY *Vice President*
DEBORAH HAMPTON *Vice President*
DANIEL ZOHNY *Treasurer*
SERGIO BARRAGAN *Secretary*
ANTHONY DREYER *Counsel*
ETIENNE SANZ DE ACEDO *Chief Executive Officer*

The Trademark Reporter Committee

EDITORIAL BOARD

EDITOR-IN-CHIEF, CHAIR

RAFFI V. ZEROUNIAN

STAFF EDITOR-IN-CHIEF

WILLARD KNOX

VICE CHAIR

PAMELA CHESTEK

Senior Editors

ANDREW GRAY
KAREN ELBURG
LESLEY GROSSBERG

VERENA VON BOMHARD
DÉSIRÉE FIELDS
BARRY HORWITZ

Director of Legal Resources

LIZ HANELLIN

Senior Staff Editor

BEVERLY HARRIS

Senior Staff Editor

ELIZABETH VENTURO

Editor-in-Chief, Practice Guides and Legal Resources

ROSEMARY DESTEFANO

Composer

BARBARA MOTTER

Editors

TARA ALLSTUN
ABDURRAHIM AYAZ
SARA BAUER
DANIEL BERESKIN
WILLIAM BORCHARD
STEPHANIE BUNTIN
JEANNETTE CARMADELLA
JOSEPH CARRAFIELLO
TED DAVIS
MICHAEL DENNISTON
MOHAN DEWAN
TUE DO
MATTHEW EZELL
CHRISTINE FELLER
MATHILDE FLORENSON
GABRIELE FOGNER
ALEX GARENS
STUART GREEN
STACY GROSSMAN
JORDI GÜELL
WINSLOW HALL

JOHN HEINBOCKEL
GANG HU
BRUCE ISAACSON
MANAVI JAIN
SAUMYA KAPOOR
INGRIDA KARINA-BERZINA
ELISABETH KASZNAR FEKETE
CHIAKI KATO
SUSAN KERI
SABRINA LARSON
SCOTT LEBSON
BARBARA LEITAO-JOHNSON
QIANG MA
SONAL MADAN
ILARIA MAGGIONI
MICHAEL MAIER
JAMES MCALLISTER
J. THOMAS MCCARTHY
CATHERINE MITROS
ERIC MORAN

SHANA OLSON
RICCARDO PEROTTI
LUIS HENRIQUE PORANGABA
UDAYVIR RANA
BRANDON RESS
SUSAN RUSSELL
MARTIN SCHMIDT
TOM SCOURFIELD
JENNIFER SICKLER
GIULIO ENRICO SIRONI
SCOTT SLAVICK
MEGAN SNEED
RANDY SPRINGER
DONALD THOMPSON
KHALA TURNER
CRISTINA VALCAUAN
MARTIN VIEFHUES
JOHN L. WELCH
JASON WHITNEY
ALICJA ZALEWSKA
NANCY ZHANG

Advisory Board

MILES ALEXANDER
WILLIAM BORCHARD
LANNING BRYER
JESSICA CARDON
SANDRA EDELMAN
ANTHONY FLETCHER
ROBERT KUNSTADT

THEODORE MAX
KATHLEEN MCCARTHY
GLENN MITCHELL
JONATHAN MOSKIN
VINCENT PALLADINO
JOHN PEGRAM

ROBERT RASKOPF
PASQUALE RAZZANO
SUSAN REISS
PIER LUIGI RONCAGLIA
HOWARD SHIRE
JERRE SWANN, SR.
STEVEN WEINBERG

The views expressed in *The Trademark Reporter* (TMR) do not necessarily reflect those of the International Trademark Association (INTA). To fulfill its mission of delivering cutting-edge scholarship on trademarks, brands, and related intellectual property to its readers, the TMR sources content reflecting a diversity of viewpoints; the views expressed in any given article, commentary, or book review are those of the individual authors.

The TMR (ISSN 0041-056X) is published electronically six times a year by INTA, 675 Third Avenue, New York, NY 10017-5704 USA. INTA, the INTA logo, INTERNATIONAL TRADEMARK ASSOCIATION, 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.

The Trademark Reporter®

(USPS 636-080)

Vol. 113

September–October, 2023

No. 5

TABLE OF CONTENTS

ARTICLES

A History of the Evolution of Likelihood of Confusion Methodologies Jerre B. Swann.....	723
Why Ask Why? A Critical Assessment of an Historical Survey Artifact R. Charles Henn Jr.....	772

The Trademark Reporter®

A HISTORY OF THE EVOLUTION OF LIKELIHOOD OF CONFUSION METHODOLOGIES

*By Jerre B. Swann**

TABLE OF CONTENTS

I.	Introduction	724
II.	The History of the Eveready Survey Format.....	725
	A. Subsequent Applications of the Eveready Format	744
	B. Variants of the Eveready Format	748
III.	The History of the Squirt Format	751
	A. Subsequent Applications of the Squirt Format	761
	B. Variants of the Squirt Format	764
IV.	Other Methodologies.....	765
V.	Rejected Methodologies.....	767
VI.	Controls	767
VII.	Conclusion	770

* Retired Partner, Kilpatrick Townsend & Stockton, Atlanta, Georgia, Member, International Trademark Association; former Editor-in-Chief, *The Trademark Reporter*.

I. INTRODUCTION

It is axiomatic that a likelihood of appreciable trademark confusion cannot arise absent appreciable opportunities for consumers contextually to compare the marks at issue. There are two basic methodologies to assess whether two marks are likely to be compared and confused by consumers in a real-world setting: (i) the unaided Eveready format¹ where trademark information as to the senior mark sufficiently exists in memory so as to be cued by a similar junior mark² and ***compared in the mind***; and (ii) the aided Squirt format³ for marks that are less (if at all) accessible in memory, but are competitively proximate so as to be ***compared in the marketplace***.⁴ For marks neither accessible in memory nor competitively proximate, the opportunity for their comparison (and for confusion) is effectively nonexistent.⁵

I have authored or co-authored three substantive articles⁶ and two book chapters⁷ on likelihood of confusion surveys in trademark cases. In my first article and chapter, I (i) over-emphasized “accessibility”⁸ of the senior brand in memory as a predicate for effective use of the Eveready format; and (ii) stressed only physical or temporal proximity⁹ as a predicate for the use of the Squirt format. In subsequent articles and in the Second Edition, I have thus necessarily engaged in extensive revisions.

¹ Union Carbide Corp. v. Ever-Ready, Inc., 531 F.2d 366, 385-88 (7th Cir. 1976).

² For reverse confusion, respondents are asked who puts out the senior use to assess whether the junior use has saturated the market and effectively appropriated the senior’s value.

³ SquirtCo v. Seven-Up Co., 628 F.2d 1086, 1089 n.4, 1091 (8th Cir. 1980).

⁴ Theoretically, the Squirt format can be used as to any marks that are competitively proximate so as to be compared in the marketplace, irrespective of the senior mark’s accessibility in memory. Given, however, the judicial preference for the Eveready format discussed *infra*, I am not aware of instances where the Squirt format has been deployed as to a mark that is readily accessible in memory.

⁵ See Jerre B. Swann, *Eveready and Squirt—Cognitively Updated*, 106 TMR 727, 728 (2016).

⁶ Jerre B. Swann, *Likelihood of Confusion Studies and the Straitened Scope of Squirt*, 98 TMR 739 (2008) (hereinafter *Straitened Scope of Squirt*); Jerre B. Swann, *Eveready and Squirt—Cognitively Updated*, 106 TMR 727 (2016) (hereinafter *Cognitively Updated*); Jerre B. Swann & R. Charles Henn Jr., *Likelihood of Confusion Surveys: The Ever-Constant Eveready Format; The Ever-Evolving Squirt Format*, 109 TMR 671 (2019) (hereinafter *Ever-Constant Eveready*).

⁷ Jerre B. Swann, *Likelihood of Confusion*, in Trademark and Deceptive Advertising Surveys 53-78 (Shari Seidman Diamond & Jerre B. Swann eds., 1st ed. 2012) (hereinafter First Edition); Jerre B. Swann, *Likelihood-of-Confusion Surveys*, in Trademark and Deceptive Advertising Surveys 59-78 (Shari Seidman Diamond & Jerre B. Swann eds., 2d ed. 2022) (hereinafter Second Edition).

⁸ *Straitened Scope of Squirt*, *supra* note 6, at 739, 745 (effectively adopting the then prevailing “top-of-mind” standard articulated by Phyllis J. Welter in Trademark Surveys § 24.03[1][c] (1999)). See *infra* note 61.

⁹ First Edition, *supra* note 7, at 68, 70.

In *Cognitively Updated*, for example, I observed that the similarity of marks is the equal of their accessibility for Eveready format purposes¹⁰ and that there can thus be no fixed level (e.g., top-of-mind) for either as critical to the execution of an Eveready study. In the Second Edition, I added salience, familiarity, uniqueness and aided awareness as alternative accessibility considerations for Eveready purposes.¹¹ In *Ever-Constant Eveready*, I stressed that **competitive** proximity, as opposed to mere physical or temporal proximity (e.g., shaving gel and razor blades), is critical for Squirt format purposes.¹²

It has been an interesting journey and one which I only now appreciate can never truly end. In this article, I will amplify my Eveready and Squirt historical efforts, briefly trace the subsequent uses of each format, explore variants and consider other methodologies that may merit future use. I will not conclude, however, with any sense that my likelihood of confusion musings have finally ended.

II. THE HISTORY OF THE EVEREADY SURVEY FORMAT

In *Union Carbide*, the Seventh Circuit endorsed the Eveready format, explicitly reversing the district court.¹³ Plaintiff's predecessor had (i) adopted EVER READY for electrical appliances in 1898 and (ii) changed to EVEREADY in 1901. Plaintiff (iii) owned five registrations for EVEREADY, alone and combined with other words and designs; (iv) had filed § 1058 (for continued validity) and § 1065 (for incontestability) affidavits for each; (v) had continuously marketed and extensively advertised (for that time) a line of batteries, flashlights and bulbs under the mark; and (vi) enjoyed sales in the 1960s exceeding \$100 million dollars per year.¹⁴ In 1969, defendant began importing Ever-Ready labeled lamp mini bulbs and high intensity lamps.¹⁵

To support an inference as to a likelihood of confusion, plaintiff offered consumer surveys as to both the mini bulbs and the lamps, asking as to the bulbs:

Question 1: Who do you think puts out these mini-bulbs?
[displaying a bulb blister pack labeled Ever-Ready].

Question 2: What makes you think so? Question 3a: Have

¹⁰ *Cognitively Updated*, *supra* note 6, at 735.

¹¹ Second Edition, *supra* note 7, at 62-63.

¹² *Ever-Constant Eveready*, *supra* note 6, at 677.

¹³ *Union Carbide Corp. v. Ever-Ready, Inc.*, 531 F.2d 366, 385-88 (7th Cir. 1976).

¹⁴ *Union Carbide Corp. v. Ever-Ready, Inc.*, 392 F. Supp. 280, 283 (N.D. Ill. 1975), *rev'd*, 531 F.2d 366 (7th Cir. 1976).

¹⁵ 531 F.2d at 371.

you seen or heard of any advertising by the concern which you think puts out these mini-bulbs? Question 3b: Please specify where, what type and features you recall. Question 4: Please name any other products put out by the same concern which you think puts out these mini-bulbs?¹⁶

As to the lamps, Question 3 (a and b) as to advertising was eliminated:

Question 1: Who do you think puts out the lamp shown here [labeled Ever-Ready on the base]?¹⁷ Question 2: What makes you think so? Question 3: Please name any other products put out by the same concern which you think puts out the lamps shown here.¹⁸

Union Carbide was a classic example of the anonymous source doctrine. While the Seventh Circuit noted that “[everyone] living in our society, . . . [is] thoroughly acquainted with the EVEREADY mark,”¹⁹ only 0.6% of respondents in the lamp study answered that Union Carbide “put out” the Ever-Ready lamp;²⁰ 54.6% answered, however, that batteries were put out by the same concern²¹—which the Seventh Circuit held was evidence of a consumer association of the marks “substantially higher [than had been] sufficient in other cases to support . . . an inference that confusion [was] likely.”²²

In the half century since the Seventh Circuit’s decision, the lamp study has become the gold standard of courts²³ and the trademark office.²⁴ It could easily have been otherwise. In the district court, both the lamp and the bulb studies were decimated on cross examination of the plaintiff’s survey expert, and the Seventh Circuit failed to address a flaw in the surveys that, in my opinion, substantially inflated the percentage of batteries responses.

On cross examination, Union Carbide’s survey expert first agreed that the questionnaire in *General Motors Corp. v. Cadillac*

¹⁶ 392 F. Supp. at 292.

¹⁷ “Puts out” is archaic. In studies executed by Dr. Gerald L. Ford and Dr. AnnaBelle Sartore, my Eveready format mentors, the question was reworded for clarity: “Who makes or puts out . . . ?”

¹⁸ 392 F. Supp. at 292.

¹⁹ 531 F.2d at 381.

²⁰ *Id.* at 385-86.

²¹ *Id.*

²² *Id.* at 386. The “batteries” responses also supported a finding of secondary meaning for EVEREADY. *Id.* at 381.

²³ J. Thomas McCarthy, *McCarthy on Trademarks and Unfair Competition* § 32:174 (5th ed. 2023) (citing *Cognitively Updated*, *supra* note 6). In the Second Edition, *supra* note 7, and in this history, I correct the top-of-mind frailty that McCarthy notes in my early Eveready discussions.

²⁴ Deborah Jay, *He Who Steals My Good Name: Likelihood of Confusion Surveys in TTAB Proceedings*, 104 TMR 1141, 1149-50 (2014).

*Marine & Boat Co.*²⁵ with the question, “Will you please name anything else you think is put out by [Cadillac Boats]?”²⁶—**substantively identical to Question 3 in the lamp study and Question 4 in the mini bulb study**—was *leading* and could possibly create “bias” and “error.”²⁷ Dominant brands like CADILLAC in the luxury automobile segment and EVEREADY as to batteries: (i) exist in memory as elaborate schemas (“clusters” of information) with source identifying/reputational nodes at their core, *strongly linked to the products or services in connection with which they are used*; and (ii) a question, therefore, as to anything else put out by CADILLAC Boats could (and did) *lead to* “cars” responses,²⁸ just as a question as to “other products put out by the same concern which you think puts out” [Ever-Ready labelled mini bulbs or an Ever-Ready labeled lamp] substantially **triggered**, in my opinion, “batteries” responses. In the district court’s view, the similarities between the CADILLAC and EVEREADY questionnaires should have evoked testimony from plaintiff’s expert that the bulb and lamp studies could likewise lead to bias and error.²⁹

Second, and further as to bias, plaintiff’s expert testified that Question 3 was removed from the lamp study because “there was no advertising performed by the defendant on [lamps]”:

Question: Well, what did [plaintiff’s counsel] say to you on that specifically?

Answer: They said that . . . when they asked me to prepare a questionnaire for the lamp study, I prepared an identical questionnaire for both studies, and they said, well . . . the defendant has not done any advertising. Do you think we should ask that particular question? I said no.

Question: And why did you think that that question should not be included if in fact the defendant had done no advertising of the lamp?

Answer: Because if we insert it there, I think . . . we could be accused of showing a bias for the Union Carbide Company.

Question: Well, how could you be accused of showing a bias for the Union Carbide Company by including Question [3a]

²⁵ 226 F. Supp. 716 (W.D. Mich. 1964).

²⁶ *Id.* at 737.

²⁷ *Union Carbide Corp. v. Ever-Ready, Inc.*, 392 F. Supp. 280, 293 (N.D. Ill. 1975).

²⁸ 226 F. Supp. at 736.

²⁹ 392 F. Supp. at 293-94. On redirect, plaintiff’s expert attempted to limit his testimony as to leadingness to what the judge in *Cadillac* had concluded as to the format in that case, but the district court expressly rejected the effort. *Id.* at 293 n.25. The defendant implied on appeal that the expert had been coached during a recess. *Union Carbide Corp. v. Ever-Ready, Inc.*, 531 F.2d 366, 387 (7th Cir. 1976).

if the defendant did no advertising of its product being surveyed.

Answer: Because if we asked that question of a respondent and the defendant did not use it, how could they possibly answer in terms of the defendants' product?

Question: And, therefore, the only advertising that would be likely to be called to mind would be Union Carbide's advertising?

Answer: That would be correct.³⁰

The district court noted that (i) "Carbide spent in excess of \$50 million for advertising and promoting the sales of its EVEREADY products between 1943 and 1973"; (ii) Ever-Ready "does not advertise its mini-bulbs" or its lamps; (iii) "[t]he same [bias] rationale is applicable to [both studies];" and (iv) "[s]uccinctly, the bulb and lamp surveys do not establish the requisite likelihood of confusion between the term 'Ever-Ready' and the mark EVEREADY."³¹

As to leadingness, the Seventh Circuit noted that "[w]hy this characterization was justified is not clearly explained"³² and observed that the mechanics of the *Cadillac* study were "sloppy":

The sample [in *Cadillac*] was of only about 150 persons. Many had no knowledge of boats and were not "purchasers." The questioning was conducted by two college students, and the tabulations were held to be "neither accurate nor truly reflective."³³

The Seventh Circuit, however, did not address the district court's tacit observation that the "Ever-Ready" label on the mini bulbs and lamps evoked the EVEREADY schema and triggered an unknown percentage of "batteries" responses as to other products put out by the same concern.

If the Seventh Circuit had concurred in the district court's rejection of the bulb and lamp studies for failure to utilize a "control" for **brand dominance** (as is now commonly done),³⁴ it would have made a lasting (and then much needed) contribution to the science of likelihood of confusion studies.³⁵ In that event, however, the

³⁰ 392 F. Supp. at 293-94.

³¹ *Id.*

³² 531 F.2d at 386.

³³ *Id.*

³⁴ *Straitened Scope of Squirt*, *supra* note 6, at 746; *Ever-Constant Eveready*, *supra* note 6, at 672; Second Edition, *supra* note 7, at 61; *see also* AHP Subsidiary Holding Co. v. Stewart Hale Co., 1 F.3d 611, 613 (7th Cir.1993); *Combe Inc. v. Dr. August Wolff GmbH & Co. KG Arzneimittel*, 382 F. Supp. 3d 429 *passim* (E.D. Va. 2019); *Sara Lee Corp. v. Goldstone Hosiery Co.*, 2001 WL 253616, at *8-10 (T.T.A.B. Feb. 22, 2001).

³⁵ For decades, experts argued that because open-ended Eveready questions "require the respondent to . . . express an answer in his or her own words," Shari Seidman Diamond,

heading of this section, *supra* Part II, would not likely be *The History of the Eveready Survey Format*.

As for the district court's holding that the Eveready format was leading, "closed-ended" Squirt questions as to only two stimuli have often been deemed leading (or overly suggestive),³⁶ and a question as to "anything else . . . put out by [Cadillac Boats]" did *cognitively* trigger (lead to) "cars" responses. While open-ended Eveready questions are not leading in any sense of that term, and the *Cadillac* questions "on their face [were] not leading,"³⁷ the elaborate CADILLAC schema and dominance of the CADILLAC brand in the luxury automotive segment could (and did) trigger an unknown quantum of "cars" responses as other products put out by CADILLAC boats. Likewise, consumers' thorough acquaintance with the EVEREADY mark cognitively suggested a "batteries" response in *Union Carbide* as other products put out by the maker of an Ever-Ready labelled lamp. The district court had a valid question as to deployment of the Eveready format in the context of a dominant mark. The Seventh Circuit did not articulate an appropriate response.³⁸

Reference Guide on Survey Research, in *Reference Manual on Scientific Evidence* 391 (3d ed. 2011), they were not as suggestive as closed-ended questions with response options and did not require controls. Dr. Diamond is, however, almost single-handedly responsible for the current requirement of controls in all likelihood of confusion surveys where causation is at issue, and if the Seventh Circuit had insisted on a control for the public's thorough familiarity with EVEREADY batteries, it would have substantially shortened the Eveready format's ascension to true gold standard status.

³⁶ See *Universal City Studios, Inc. v. Nintendo Co.*, 746 F.2d 112, 118 (2d Cir. 1984) ("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?" was "an obvious leading question in that it suggested its own answer."); *Riviana Foods, Inc. v. Societe Des Produits Nestle S.A.*, 33 U.S.P.Q.2d (BL) 1669, 1671 (S.D. Tex. 1994) ("Do you think the weight loss product 'Sweet Success' and 'Success Rice' are more likely made by the same company or more likely made by different companies?" was "leading" because it followed other questions that may have "creat[ed] an association between the two products where none may have existed previously."); *Hornady Mfg. Co. v. Doubletap, Inc.*, 746 F.3d 995, 1006 (10th Cir. 2014) ("By suggesting the possibility that [Defendant] might be connected with another brand, and limiting the candidates to [Plaintiff] and Federal Premium, the [Defendant's] Survey risked sowing confusion between [Plaintiff] and TAP when none would have arisen otherwise."); *Beneficial Corp. v. Beneficial Cap. Corp.*, 529 F. Supp. 445, 450 (S.D.N.Y. 1982) ("Do you think that there may or may not be a business connection between Beneficial Capital Corp. and the Beneficial Finance System Companies?" was "not well suited to eliciting an uninfluenced reaction."); *Kargo Glob., Inc. v. Advance Mag. Publishers, Inc.*, No. 06 Civ. 550(JFK), 2007 WL 2258688, at *5, *8 (S.D.N.Y. Aug. 6, 2007) ("[D]o you think these ads from Premiere Magazine and for the cell phone service you just saw . . . ? [1.] do come from the same company . . . [2.] do not come from the [same] company . . . [3.] Don't know" was suggestive because "the mere putting of [the] question create[d] the impression of a relationship."). The use of only two stimuli is particularly problematic where the marks at issue are not competitively proximate so as to create artificial proximity and/or produce demand effects.

³⁷ 531 F.2d at 386-87.

³⁸ In fairness, I do not believe that brand association networks were then commonly discussed in survey literature.

As for the bias created by the advertising question in the mini bulb study, the Seventh Circuit acknowledged that (given the disparity in the parties' expenditures) there was a problem,³⁹ but noted the similarity of the lamp and mini bulb study results and opined that "we believe the likelihood of substantial bias as a result of the advertising question is small."⁴⁰ I have argued that any evidence of bias is potentially fatal (because its impact can rarely be quantified), and Question 3 likely exacerbated the brand dominance of EVEREADY that plagued the survey as a whole: (i) as plaintiff's expert testified, respondents could have *only* "seen" EVEREADY advertising and had reinforced in their memory its schema (with its link to batteries); and (ii) respondents were given yet another exposure to the aurally identical "Ever-Ready" mark emblazoned on the bulb blister packs and lamp bases which also clearly triggered the EVEREADY schema (and its strong batteries component).

Given its (correct) view that: (i) incontestability negated the defendant's descriptiveness defense; (ii) there was "overwhelming" evidence of secondary meaning; (iii) consumers were "thoroughly acquainted" with the mark EVEREADY and its "battery-operated" products (531 F.2d at 381); (iv) there was evidence of actual confusion; and (v) the marks were legally and aurally identical, it is my view that the Seventh Circuit, notwithstanding its glowing verbiage as to the impact of the survey,⁴¹ would have reversed the district court even had it agreed that the survey was biased and leading (or did not control for brand dominance). I do not know of any case where a junior user's essentially identical adoption of a dominant, incontestable mark, overwhelmingly known to consumers and used on a related category of goods has not been held likely to cause confusion. The Seventh Circuit noted that defendant's counsel in his opening argued (and *admitted*) that "**[w]e don't sell batteries, and that's what everybody thinks of when you mention the name EVEREADY.**"⁴² Plaintiff's survey was *surplusage*.

It is not necessary, however, to debate the correctness of the district court's analysis. First, Question 3 as to advertising in the bulb study was not included in the lamp study that has become the Eveready format as we know it today. Second, batteries and high intensity lamps were "not . . . competing goods"⁴³ and the marks were essentially identical. The "other products" question was thus

³⁹ 531 F.2d at 388.

⁴⁰ *Id.*

⁴¹ 531 F.2d at 388.

⁴² 531 F.2d at 381 (emphasis added). "The strongest brands . . . own a place in the consumer's mind, and when they are mentioned almost everyone thinks of the same things." Scott M. Davis, Brand Asset Mgmt. 3 (2002). Defense counsel's admission thus placed Eveready in an elite category of marks.

⁴³ Union Carbide Corp. v. Ever-Ready, Inc., 392 F. Supp. 280, 289 (N.D. Ill. 1975).

likely deemed necessary to demonstrate the “reach” of the battery brand to lamps and to differentiate between respondents who in answer to Question 1 merely played back the “Ever-Ready” label on the lamp, from those who believed that the lamp was put out by the battery company.⁴⁴ Today, (i) the format is often used with respect to competing or similar products, (ii) “reach” is rarely an issue,⁴⁵ (iii) the junior mark is rarely visually or aurally identical to a dominant senior, and (iv) the “other products” question is often not needed.⁴⁶

Irrespective, therefore, of the Eveready format’s somewhat clouded origins, I have most recently repeated that:

A readily accessible “familiar” mark in memory [that will likely produce an appreciable percentage of senior brand responses to a makes or puts out question] can thus be [cognitively] conceptualized as residing near the surface of a *consumer’s memory* and potentially *reaching* to any market into which the consumer travels (irrespective of whether the senior user’s product or service is sold therein). The proposition that the Eveready format measures reasonably frequent opportunities for mental comparisons is supported by two considerations: (i) it produces evidence in support of an inference as to a likelihood of confusion only where a senior mark is sufficiently accessible in memory to be triggered by a similar junior mark; and (ii) it is conducted among consumers in the category of (and thus likely to encounter) the junior mark. As such, an Eveready study is a *reliable measure of real-world accessibility and similarity, i.e., of whether the senior use is sufficiently close to the surface of memory, and the junior use is sufficiently*

⁴⁴ Second Edition, *supra* note 7, at 61.

⁴⁵ See *Kenner Parker Toys Inc. v. Rose Art Indus., Inc.*, 963 F.2d 350, 353 (Fed. Cir. 1992) (“A strong mark . . . casts a long shadow . . .”). An Eveready survey can be used today by a defendant to suggest a brand shadow’s limits, as well as by a plaintiff to suggest its full length or reach.

⁴⁶ Particularly now that controls address brand dominance, the “other products” question might still be necessary and appropriate in another similar-marks/different-products case, particularly if there is any question as to secondary meaning. If, e.g., another company adopted NIKE for bottled water, respondents who (in appreciable numbers) answered “shoes” as other products that it put out would reflect that the shadow of the NIKE mark reached to bottled water and that NIKE is substantially associated with its current owner. See *Tiffany & Broadway, Inc. v. Comm’r of Pats. & Trademarks*, 167 F. Supp. 2d 949, 954-55 (S.D. Tex. 2001) (26% of respondents indicated a relationship between Tiffany shoes and Tiffany the jeweler); *Gap, Inc. v. G.A.P. Adventures, Inc.*, 100 U.S.P.Q.2d (BL) 1417, 1425-26 (S.D.N.Y. 2011) (40% of respondents confused into thinking that G.A.P Adventures offered clothing or was operated/authorized by Gap, but when defendant’s full name was used (Great Adventure People Adventures) as a control, zero confusion was evident).

*similar to the senior use so that . . . the former will likely be activated (recalled) by a monadic exposure to the latter.*⁴⁷

From a cognitive standpoint, *pattern recognition* or *matching* is at the *heart* of the Eveready format:

1. “When a sufficient number of features represented in the incoming [brand] information **match[es]** the pattern of features of a pre-existing [brand] network [in memory], we tend to fill in the details [‘Wheel of Fortune’-like] and interpret the object as an exemplar of that [brand] network.”⁴⁸
2. When shown an allegedly infringing junior mark and asked who makes or puts it out, respondents search their memory **introspect** and identify the mark “based on its **similarity** to what they already know.”⁴⁹
3. “[W]e have in our head a remarkably powerful computer, . . . able [rapidly] to represent the structure of our world by various types of associative links in a vast network of various types of ideas.”⁵⁰
4. “When stimulus information offers a sufficient [**fit**] to a schema possessed by the perceiver, the schema is called up from memory and used . . . to guide inferences.”⁵¹

As such, the Eveready format *may* be self-validating: i.e., (i) reviews of “why do you say that?” answers can reveal that senior mark responses to the “who makes or puts out” question have likely occurred (ii) because stored knowledge of the senior mark is “**accessib[le]**” in respondents’ memories and (iii) there is a “**fit**” between that knowledge and the junior stimulus.⁵² In *Facebook, Inc. v. Think Computer Corp.*, for example, the Board noted that the responses to the follow-up question—why do you say that—made “specific reference to the similarities between the names FACEMAIL and FACEBOOK.”⁵³ I.e., FACEMAIL “stimulus

⁴⁷ Second Edition, *supra* note 7, at 64 (quoting *Ever-Constant Eveready*, *supra* note 6, at 672-73).

⁴⁸ Jacob Jacoby, *The Psychological Foundations of Trademark Law, Secondary Meaning, Genericism, Fame, Confusion and Dilution*, 91 TMR 1013, 1035, 1037 (2001) (emphasis added) (compulsory reading for an understanding of cognitive concepts impacting likelihood of confusion assessments).

⁴⁹ Wayne D. Hoyer & Deborah J. MacInnis, *Consumer Behavior* 102, 108, 115-16 (3d ed. 2004) (emphasis added).

⁵⁰ Daniel Kahneman, *Thinking, Fast and Slow* 71 (2011).

⁵¹ D.E. Carlston & E.R. Smith, *Principles of Mental Representation*, in *Social Psychology: Handbook of Basic Principles* 196 (E.T. Higgins & A.W. Kruglanski eds., 1996).

⁵² See E.T. Higgins, *Knowledge Activation: Accessibility, Applicability, and Salience*, in *Social Psychology: Handbook of Basic Principles* 135 (E.T. Higgins & A.W. Kruglanski eds., 1996).

⁵³ 2013 WL 4397052, at *15 (T.T.A.B. July 23, 2013). As other examples of fit/pattern

information” was similar to/fit the FACEBOOK schema possessed in memory “by the perceiver, [so that the FACEBOOK schema was] called up from memory and used . . . to guide inferences” as to FACEMAIL.⁵⁴

Consumers, of course, “rarely . . . consider all [stimulus] features,”⁵⁵ and because respondents can sometimes have difficulty accessing their higher order processes,⁵⁶ pattern matching is (most) often not evident. “Why” answers, moreover, can never serve as controls to establish causation⁵⁷ and may, indeed, produce “pseudo-reasons [because] respondents who are questioned about reasons for an answer will search for a plausible explanation that may or may not be the reason for their earlier response.”⁵⁸ In my experience,

matching, *see* Audemars Piquet Holding S.A. v. Swiss Watch Int’l, Inc., 46 F. Supp. 3d 255, 277 (S.D.N.Y. 2014) (“Despite the fact that the watches in the study had a Swiss Legend logo on their faces, those who identified them as Audemars watches nonetheless listed the aspects of the Royal Oak trade dress design as the reason, including the octagonal shape of the watch (23%), the style, design or shape of the watch dial (20%), [and] the octagonal bezel and shape or look (15%)”), *rev’d in part on other grounds*, No. 12 Civ. 5423(LAP), 2015 WL 150756 (S.D.N.Y. Jan. 12, 2015); PepsiCo, Inc. v. Pirincci, 2012 WL 2930650, at *8 (T.T.A.B. June 25, 2012) (“[A]ll of the 125 confused respondents gave answers indicating that the word ‘Dew’ was the reason for their confusion [of the CAN DEW and MOUNTAIN DEW marks.]”); H-D Mich., Inc. v. Top Quality Serv., Inc., 496 F.3d 755, 758 (7th Cir. 2007) (why answers suggested that confusion between HOGS ON THE HIGH SEAS and Harley-Davidson’s HOG and H.O.G. marks stemmed from the defendant’s “use of the word ‘hog’ and [that] word’s association with the Harley Owners Group [H.O.G.]”); Fiji Water Co. v. Fiji Min. Water USA, LLC, 741 F. Supp. 2d 1165, 1179 (C.D. Cal. 2010) (many respondents’ reasons for thinking defendants’ product was associated with plaintiffs were the elements of plaintiffs’ trade dress); Starbucks U.S. Brands, LLC v. Ruben, 78 U.S.P.Q.2d (BL) 1741, at *39 (T.T.A.B. 2006) (answers to the “What makes you think so” question “persuade us that the respondents were not merely guessing”).

⁵⁴ *See* Carlston & Smith, *supra* note 51 and accompanying text; *see also* Gen. Motors Co. v. Urban Gorilla, LLC, No. 2:06-CV-00133 BSJ, 2010 WL 5395065, at *18 (D. Utah Dec. 27, 2010) (“The respondents in this case repeatedly said the Urban Gorilla was a ‘Hummer’ and when asked why, the oft recurring response was ‘because it looks like one.’ These responses strongly suggest, if not prove, actual confusion.”). The look of the Hummer was close to the surface of memory and was “oft” cued by the very similar look of the Urban Gorilla.

⁵⁵ Jacoby, *supra* note 48, at 1035.

⁵⁶ Richard E. Nisbett & Timothy DeCamp Wilson, *Telling More than We Can Know: Verbal Reports on Mental Processes*, 84 *Psych. Rev.* 231, 231-232 (1977).

⁵⁷ With the advent of experimental designs, Dr. Diamond is of the opinion (and I agree) that “why” questions are no longer necessary. Because, however, (i) some courts place controlling value on why answers, *see supra* note 53; (ii) others like to analyze the clarifying information that why questions can sometimes provide, *see* Cumberland Packing Corp. v. Monsanto Co., 32 F. Supp. 2d 561, 572-76 (E.D.N.Y. 1999); (iii) a few courts have rejected studies without why questions, *e.g.*, Pep Boys Manny, Moe & Jack of Cal. v. Goodyear Tire & Rubber Co., No. 01-CV-5614, 2002 WL 524001, at *10 (E.D. Pa. Apr. 5, 2002), and (iv) why answers sometimes can be helpful to me in interpreting a survey, I will continue to insist on their inclusion in studies I commission. *See Straitened Scope of Squirrt, supra* note 6, at 741 n.13.

⁵⁸ Shari Seidman Diamond, *Control Foundations: Rationales and Approaches*, in First Edition, *supra* note 7, at 211.

however, and as counsel for Adidas, it was my view that a *consistent pattern*, across multiple questionnaires and thousands of respondents, of 35% or more of respondents stating that Adidas was the maker of a two/four stripe athleisure shoe,⁵⁹ with many giving “stripes” as the reason for their answer, reflected that respondents were accessing the adidas three stripe brand schema in their memory, and that pattern matching was the most plausible explanation for their response that Adidas made or put out a two/four stripe athleisure shoe, i.e., as in *Facebook, supra*, the survey results were self-validating.⁶⁰

To trigger pattern matching in the unaided Eveready format, Phyllis Welter continued to postulate in her last (1999) edition of *Trademark Surveys* that “‘top of mind’ awareness of [the senior] brand [was] required.”⁶¹ As, however, a study of the *contents of memory*, the format is extendable (i) to other categories of marks and (ii) to uses beyond likelihood of confusion.

First, an Eveready study clearly **does not assess** or **require** top-of-mind awareness of the senior mark. With, indeed, a similar junior mark as a stimulus, the format more closely resembles a partially “aided awareness” test,⁶² and commercially strong marks, particularly for high-involvement/salient goods (e.g., athletic shoes), can reach far deeper into memory.⁶³ A swoosh is likely the top-of-mind symbol in the athletic shoe category. Shoes, however, bearing two or four parallel stripes extending from the sole to the lace area evoke, as noted above, high percentages of Adidas answers to the “who makes or puts out” question (expectations often drive perception and respondents may overwrite features of stimuli to “match” their memory, converting two/four stripes to adidas’s readily accessible, familiar three).⁶⁴

⁵⁹ *E.g.*, Adidas Am., Inc. v. Payless Shoesource, Inc., 529 F. Supp. 2d 1215, 1241, 1246 (D. Or. 2007).

⁶⁰ Dr. Ford conducted the surveys as to FACEBOOK and as to two/four stripe athleisure shoes and opined in his testimony that the results in the latter were self-validating. His testimony was, of course, very helpful, but the “why” answer results were in no way required or necessary to demonstrate that his data supported an inference as to likelihood of confusion. His partner, Dr. Sartore, always agreed with me as to the inclusion of why questions, even though we appreciated that why answers can be confounding. *See supra* note 57.

⁶¹ Phyllis J. Welter, *Trademark Surveys* § 24.03[1][c] (1999). A top-of-mind brand is typically the first brand given in response to a category prompt. Readily accessible, “familiar” and unique brands may also be given in response to probing and are often sufficiently near the surface of memory to be cued by a “makes or puts out” Eveready question.

⁶² *Straitened Scope of Squirt, supra* note 6, at 745 n.40.

⁶³ *Cognitively Updated, supra* note 6, at 733.

⁶⁴ *E.g.*, Adidas Am., Inc. v. Payless Shoesource, Inc., 529 F. Supp. 2d 1215, 1241 (D. Or. 2007) (Dr. Ford’s survey showed 41% of respondents believed Adidas made or authorized defendant Payless’s four-stripe shoe); *see RE/MAX Int’l, Inc. v. Trendsetter Realty, LLC*, 655 F. Supp. 2d 679, 704-05 (S.D. Tex. 2009) (25.3% of respondents answered that RE/MAX sponsored or promoted a real estate sign with red-over-white-over-blue bars,

As I further noted in *Cognitively Updated*:

Two basic variables influence the likelihood that some stored knowledge will be activated—the *accessibility* of the stored knowledge prior to stimulus presentation [its nearness to the surface of memory] and the *fit* [perceived similarity] between the stored knowledge and the presented stimulus. . . . The greater is the overlap between the features of some stored knowledge and the attended features of a stimulus, the greater is the applicability of the knowledge to the stimulus and the greater is the likelihood that the knowledge will be activated in the presence of the stimulus.⁶⁵

Hal Poret confirms, in his groundbreaking 2019 TMR article, “An Empirical Assessment of the Eveready Survey’s Ability to Detect Significant Confusion in Cases of Senior Marks that Are Not Top-of-Mind,”⁶⁶ that (i) accessibility of a senior mark in memory and (ii) similarity of a junior mark to the senior are equally important ingredients for Eveready survey purposes, and, to iterate, there can thus be *no fixed* (e.g., top-of-mind) accessibility level as a prerequisite for the deployment of the Eveready format.⁶⁷

In *Ever-Constant Eveready*, I replaced “top-of-mind” with “readily accessible in memory” as marks likely to produce senior brand responses in an Eveready format.⁶⁸ Well known, “familiar” brands (frequently given in response, albeit not first-named, to a category prompt) are “selectively given more exposure, attention, comprehension and retention by consumers.”⁶⁹

“[R]are [unique] words[, moreover,] are more distinctively encoded [and are thus more easily recalled/remembered] than (are) common words.”⁷⁰ A “*unique* brand name and [a] cohesive brand identity are probably the most powerful pieces of information for

notwithstanding that “Trend Setter Realty” in large typeface was written across the front; with the red and blue bars removed, only 2.7%, “not statistically different from zero” and to be expected with respect to a category leader, mentioned RE/MAX as the company being promoted by the sign). In my experience with Adidas surveys, about 5% or so of respondents reacted that Nike put out two/four stripe athleisure shoes, which was a typical number for a category leader.

⁶⁵ *Cognitively Updated*, *supra* note 6, at 735 (quoting E.T. Higgins, *Knowledge Activation: Accessibility, Applicability, and Salience*, in *Social Psychology: Handbook of Basic Principles* 135 (E.T. Higgins & A.W. Kruglanski eds., 1996) (first italics supplied)).

⁶⁶ 109 TMR 935 (2019). For purposes of full disclosure, I often utilized Hal Poret for surveys and he once utilized me as an expert witness as to likelihood of confusion studies.

⁶⁷ *See id.* at 954; *see also Cognitively Updated*, *supra* note 6, at 735.

⁶⁸ *Ever-Constant Eveready*, *supra* note 6, at 672; *see also* Second Edition, *supra* note 7, at 61-64.

⁶⁹ Steve Hoeffler & Kevin Lane Keller, *The Marketing Advantages of Strong Brands*, 10 J. Brand Mgmt. 424 (2003).

⁷⁰ Second Edition, *supra* note 7, at 62 (quoting Joan Meyers-Levy, *The Influence of a Brand Name’s Association Set Size and Word Frequency on Brand Memory*, 16 J. Consumer Rsch. 197, 198 (1989)).

consumers . . . enabling [them] to efficiently organize, store, and retrieve information from memory.”⁷¹ Much in the manner of a picture of a celebrity on the cover of a magazine in a sidewalk kiosk, *unique identifiers attract attention* in an otherwise hurried (and harried) retail environment.⁷² The concept of uniqueness has thus worked its way into (i) confusion factor⁷³ and (ii) dilution analyses,⁷⁴ and the “rareness” of KIX (with its associated ease of recall/retrieval from memory) may thus partially account for the 25% confusion rate reported by Poret between KIX for cereal (with only 3% unaided awareness) reaching to a fictional KIXX for snack bars *in a different category*.⁷⁵

Recall from memory in response to an Eveready open-ended question is, of course, a more difficult cognitive exercise than is recognition in response to a close-ended question,⁷⁶ but a top-of-mind requirement for a brand to elicit an accurate and relevant percentage of confusion responses in the Eveready format is dramatic overkill. Degrees of accessibility—familiarity, salience and uniqueness—are more relevant considerations. A recent article, indeed, that extends Poret’s empirical approach to break still further ground as to the Eveready format, agrees with my view that the format is a partially aided awareness assessment, and posits that the “key measure for deciding whether a mark is sufficiently well known to use the Eveready format is aided awareness.”⁷⁷ Far, therefore, from being “ever-constant,” the Eveready format (in addition to now often including questions as to authorization,

⁷¹ Jacoby, *supra* note 48, at 1025 (emphasis added).

⁷² Wayne D. Hoyer & Deborah J. MacInnis, *Consumer Behavior* 118-19 (3d ed. 2004).

⁷³ *James Burrough Ltd. v. Sign of the Beefeater, Inc.*, 540 F.2d 266, 276 (7th Cir. 1976) (“A mark that is strong because of its fame or its uniqueness, is more likely to be remembered and more likely to be associated in the public mind with a greater breadth of products or services, than is a mark that is weak because relatively unknown or very like similar marks or very like the name of the product.”).

⁷⁴ 15 U.S.C. § 1125(c)(2)(B)(i)-(iv).

⁷⁵ Hal Poret, *An Empirical Assessment of the Eveready Survey’s Ability to Detect Significant Confusion in Cases of Senior Marks that Are Not Top-of-Mind*, 109 TMR 935, 953-54 (2019). Uniqueness may also explain why the FANTA and FILA marks, with low unaided awareness, triggered appreciable recall/confusion when cued by a similar/fictional junior.

⁷⁶ Jacob Jacoby, *Are Closed-Ended Questions Leading Questions?*, in *First Edition, supra* note 7, at 269-71 (“To probe the contents of memory, open-ended [Eveready-type] questions [as to a single stimulus] rely on *recall* whereas closed-ended [Squirt-type] questions [as to multiple stimuli] rely on *recognition*. For more than a century, extensive empirical research has demonstrated convincingly that, at any point in time, recognition produces greater amounts remembered than does recall. . . . Open-ended recall questions generally assess top-of-mind contents, while unbiased closed-ended . . . questions generally do a more thorough job of assessing what a person has stored in memory.”); see also David T. Neal, *Psychological Considerations in Designing Trademark and False Advertising Survey Questionnaires*, in *Second Edition, supra* note 7, at 273-90.

⁷⁷ Dr. Bruce Isaacson & Dr. Keith A. Botner, *When to Conduct an Eveready Survey: The Importance of Aided Awareness*, 111 TMR 693, 695, 696, 702 (2021).

approval and association)⁷⁸ has evolved from top-of-mind to an appreciable level of aided awareness to produce confusion percentages sufficient to support an inference as to a likelihood of confusion. And a defendant, of course, may suggest with an Eveready as to noncompeting marks that a plaintiff's mark is not in the mind of defendant's consumers at all so that no comparison is possible and no confusion is likely.⁷⁹ In a word, no level of awareness is a prerequisite to the Eveready format's relevant use.

For a more thorough *probe of memory*, it would be helpful in my view to elicit information with respect to a *marginal* Eveready study as to the balance of: (i) awareness (presence in memory) and (ii) similarity (perceived fit of marks) in producing (*or not*) a senior brand response to the "who makes or puts out" question. Only a 20% aided awareness level, for example, but a 10% senior brand response would suggest that the marks were perceived to be very similar by 50% of respondents and could convert an otherwise marginal study into useful evidence (similarity being the predominant factor in a likelihood of confusion assessment).⁸⁰ For an aggressively marketed brand like QUIRST in 1978,⁸¹ with its awareness level on the rise, a 50% similarity to SQUIRT perception by consumers would have portended rife confusion.⁸²

Because, in an Eveready study supporting an inference as to a likelihood of confusion, the junior mark is necessarily similar to the senior, it is impossible (without priming) to determine awareness levels of the senior mark after the "who makes or puts out" question has been addressed with respect to the similar junior use. Since "[a]wareness is one of the most fundamental measures in brand management,"⁸³ reliable information may well exist in the senior user's files or be separately adduced in connection with a showing of brand strength; and should (in my view) be reviewed in interpreting marginal Eveready survey results. An 80% awareness level with only a 10% senior company (perceived similarity/fit) response to the

⁷⁸ See, e.g., *Facebook, Inc. v. Think Comput. Corp.*, 2013 WL 4397052, at *14 (T.T.A.B. July 23, 2013).

⁷⁹ See *Rest. Lutece, Inc. v. Houbigant, Inc.*, 593 F. Supp. 588, 592-93 (D.N.J. 1984).

⁸⁰ Barton Beebe, *An Empirical Study of the Multifactor Tests for Trademark Infringement*, 94 Cal. L. Rev. 1581, 1600 (2006) ("The similarity of the marks factor is by far the most influential.").

⁸¹ *SquirtCo v. Seven-Up Co.*, 628 F.2d 1086, 1089 n.2 (8th Cir. 1980) ("At the time that SquirtCo first objected, Seven-Up had invested relatively little in the QUIRST mark. The investment had risen to over \$200,000 at the time of the preliminary hearing and to over \$2,000,000 by the end of 1978. By the entry of the permanent injunction, Seven-Up had expanded the product line to include orange, grape and fruit punch as well as lemonade and had expended over \$5,000,000.").

⁸² In *The History of the Squirt Survey Format*, *infra* Part III, I note that SquirtCo could (and likely should) have used an Eveready study had the format been developed at that time.

⁸³ Isaacson & Botner, *supra* note 77, at 697.

“who puts out” question would suggest a low fit perception and might doom a marginal Eveready result.

While it is my opinion that it might be instructive to elicit information in marginal cases as to the roles of awareness and similarity in producing (or not) senior user responses to a “who puts out” question, I do not believe, for two reasons, that a likelihood of confusion assessment should ever be made—as one survey expert has suggested to me in a proposed article that he asked me to review (but that to my knowledge remains unpublished)—based on a subset of respondents who have aided awareness of a mark:

1. As I note *supra*, awareness could not be computed by a back-end question without priming after the “who puts out” question as to a similar mark has been asked.
2. If, moreover, data exists as to say, a 10% level of aided awareness, and 10% of the respondents give a senior user response to the who makes or puts out question, the survey supports an inference that only a *de minimis* 1% of consumers are likely to be confused.

A real-world likelihood of confusion assessment can only be made in an overall context. Any subset calculation would verge on an attempt to test only one’s own customers which has been appropriately rejected too often to require citation.

As for evidentiary uses of the format, I have long insisted that an Eveready is the *only appropriate test* of likelihood of confusion where marks are *not competitively proximate* so as to be compared in the real-world marketplace;⁸⁴ and, as to non-competing marks, it matters not for Eveready deployment purposes whether the senior mark is top of mind, commercially strong, familiar or *inaccessible* in memory. Indeed, to iterate, by showing that a senior (non-competing) mark is not sufficiently in memory to be triggered by a similar junior use, an Eveready study may reflect that no similarity

⁸⁴ *Cognitively Updated*, *supra* note 6, at 737; see *Denimafia Inc. v. New Balance Athletic Shoe, Inc.*, No. 12 Civ. 4112(AJP), 2014 WL 814532, at *20 (S.D.N.Y. Mar. 3, 2014). Using a Squirt format as to non-competing marks inevitably places the marks in *artificial proximity*. See *Kargo Glob., Inc. v. Advance Mag. Publishers, Inc.*, No. 06 Civ. 550(JFK), 2007 WL 2258688, at *8-9 (S.D.N.Y. Aug. 6, 2007); *Hi-Tech Pharms. Inc. v. Dynamic Sports Nutrition, LLC*, No. 1:16-cv-949-MLB, 2021 WL 2185699, at *18 (N.D. Ga. May 28, 2021) (“In general, a Squirt survey is appropriate where the senior mark is not well known, and the marks often appear side by side in the marketplace. . . . But where the products at issue ‘are not sold in the same stores or, for the most part, on the same websites, such a format may over-estimate confusion by forcing consumers to consider the marks in close proximity in a way they would not in the marketplace.”); *Elevate Fed. Credit Union v. Elevations Credit Union*, No. 1:20-CV-00028-DAK-JCB, 2022 WL 798901, at *31 (D. Utah Mar. 16, 2022) (“[The *Squirt* survey] format requires that [the survey expert] carefully replicate the real-world . . . marketplace, which is done by having the survey reflect ‘a significant number of real world situations in which both marks are likely to be seen in the marketplace side-by-side.’” (citing McCarthy, *supra* note 23, § 32:174.50)).

comparison can be conducted⁸⁵ and that *no confusion is likely* in the marketplace.

As to marks that are not competitively proximate, an Eveready is, indeed, *by itself an appropriate* test for a likelihood of confusion.⁸⁶ In *any case* where a junior mark does not compete with an allegedly infringed senior mark, a defendant may dispel a likelihood of confusion with proof simply that consumers are not aware of the senior mark.⁸⁷

In *United States Hosiery Corp. v. Gap, Inc.*,⁸⁸ for example, Dr. Hans Zeisel (a father of consumer research for litigation purposes) was faced with a dispute involving U.S. Hosiery's WORKFORCE mark for socks sold in Sears and a WORKFORCE mark (sometimes with a space or hyphen after WORK) for blue jeans sold in The Gap. U.S. Hosiery had commissioned a *two stimuli Squirt* survey as to the marks and (not surprisingly) 35% of respondents answered that WORKFORCE socks and WORKFORCE blue jeans came from the same company.

Based, however, on the results of separate studies designed to measure the low (or absence of) actual awareness of the marks and the *unlikelihood* that consumers would ever encounter both products in the marketplace, Dr. Zeisel concluded on behalf of The Gap that the likelihood of confusion was not 35%, but was less than 1% among both The Gap consumers and the general population.⁸⁹ The Gap consumers generally were unlikely to encounter U.S. Hosiery's WORKFORCE socks, sold primarily at Sears, in the marketplace or have that brand for blue jeans near the surface of their memory if a chance encounter did occur. Where marks are not competitively proximate and cannot be contextually compared in the marketplace, an Eveready study, by *probing memory* as to the (low or) absence of actual awareness may alone prove that there is no likelihood of confusion either *in the mind* or *in the market*. To repeat from the *Introduction, supra* Part I, for marks neither

⁸⁵ Second Edition, *supra* note 7, at 64-65.

⁸⁶ *Cf. Cognitively Updated, supra* note 6, at 734 (positing that, for competitively proximate marks, the Eveready format by itself may not be appropriate for strong, generally-recognized, but not readily recalled marks).

⁸⁷ *See* Edison Bros. Stores, Inc. v. Cosmair, Inc., 651 F. Supp. 1547, 1556, 1560 (S.D.N.Y. 1987) (confusion unlikely where the products, clothing and perfume, were "noncompetitive" and defendant established that only 6% of respondents knew of a product sold under the NOTORIOUS name at issue); McGraw-Hill, Inc. v. Comstock Partners, Inc., 743 F. Supp. 1029, 1034-35 (S.D.N.Y. 1990) (Comstock for real time price quotations and Comstock Partners for investment consulting "are as completely unrelated as night and day" and, given the small percentage of persons surveyed that had ever heard of either company, confusion was "highly speculative"); Rest. Lutece, Inc. v. Houbigant, Inc., 593 F. Supp. 588, 592-93 (D.N.J. 1984) (no confusion likely where 39.5% of women shown boxes and containers of defendant's perfume correctly identified it and only 1.3% were aware of plaintiff's restaurant).

⁸⁸ 707 F. Supp. 800 (W.D.N.C. 1989).

⁸⁹ Hans Zeisel & David Kay, *Prove it with Figures* 169 (1997).

accessible in memory nor competitively proximate, the opportunity for their comparison (and for confusion) is effectively non-existent.

The Gap-type cases can often be resolved by a marketplace and cognitive analysis without the need for survey evidence. “[I]n cases where the two parties’ products do not share the same market or distribution channels and the products are not widely known, the likelihood of confusion is reduced [if not eliminated] by the very fact that no consumer would ever be exposed to both products.”⁹⁰ Cognitively, it is likely that the marketing files of both Sears and The Gap reflected that the respective WORKFORCE marks were low on the awareness spectrum. U.S. Hosiery’s mark was not thus near the surface of memory nor was it competitively proximate to The Gap’s brand—the Sears mark for socks thus fell into a coverage gap: there was no venue in which it could be compared to The Gap’s WORKFORCE mark for jeans and no confusion was remotely likely between the two.⁹¹

Those experts who still insist, therefore, that an Eveready is appropriate only as to top-of-mind marks miss the issue in a likelihood of confusion survey exercise. The issue is not (i) merely *accessibility* of the senior mark; rather, it is to accurately predict the results of consumer (ii) *similarity* assessments in any *mental* or *marketplace* venue where comparisons may take place. Phyllis Welter may have believed that only top of mind marks were accurate comparison predictors in an *unaided* Eveready survey context. Experience and research, however, have proven otherwise; and with non-competing brands, the Eveready format probes the degree to which the senior mark is in the mind at all and whether confusion in the marketplace is possible.

As a memory probe, the Eveready format may also be utilized to rebut exaggerated claims of strength (e.g., “of fame in connection with an ill-advised dilution add-on to a likelihood-of-confusion count”)⁹² and, as in *Union Carbide*, to evidence secondary meaning.⁹³ To iterate, the appropriate scope of the format has significantly broadened both as to (i) the range of appropriate marks for assessment and (ii) the occasions for its appropriate use in the twenty years between Welter’s last review and Poret’s 2019 *Empirical Assessment* article.⁹⁴ The “*Ever-Constant Eveready*” caption must thus be added to the list of errors that I have committed in fifteen years of writing about likelihood of confusion surveys.

⁹⁰ *Id.* at 167.

⁹¹ *Cognitively Updated*, *supra* note 6, at 744-45.

⁹² Second Edition, *supra* note 7, at 64.

⁹³ *See supra* note 22.

⁹⁴ Welter, *supra* note 61; Poret, *supra* note 75; Second Edition, *supra* note 7, at 65.

I am aware of only one significant criticism of the Eveready design, issued (oddly) in the Seventh Circuit, notwithstanding: (i) settled precedent in that and all other Circuits; and (ii) the use of control methodology in that case (*Kraft Foods Group Brands LLC v. Cracker Barrel Old Country Store, Inc.*)⁹⁵ to eliminate the principal concern expressed by the Court (that CRACKER BARREL OLD COUNTRY STORE ham and KRAFT CRACKER BARREL cheese were in sufficiently related product categories so as to promote guessing that the ham company also produced cheese).⁹⁶

There are, of course, still substantive issues. Disputes have arisen, for example, as to whether stimuli should (i) be left with respondents (referred to by critics as a reading test);⁹⁷ or (ii) be removed prior to addressing the “who puts out” question (referred to as a memory test).⁹⁸

Because relatively expensive and high involvement products like (for me) high intensity reading lamps are not removed when consumers endeavor to ascertain their provenance,⁹⁹ I likely would have left the lamps in *Union Carbide* with the respondent. I appreciate, however, that reading can be problematic¹⁰⁰ and Hans Zeisel has articulated an excellent variant for products (like many grocery items) that are fleetingly seen in the marketplace.¹⁰¹ I may well have removed the blister packs for the less expensive mini bulbs which were possibly less attended to in the market, and I would remove the stimulus in post-sale and initial interest assessments where a “glance” often triggers (or not) a similarity assessment.

⁹⁵ 735 F.3d 735 (7th Cir. 2013). In my view, there should also have been a control for the possible brand dominance of CRACKER BARREL in the cheese category as there was for PAM in the spray cooking oil segment in *AHP Subsidiary Holding Co. v. Stewart Hale Co.*, 1 F.3d 611, 613 (7th Cir. 1993), but neither the court nor the parties expressed concern in that regard.

⁹⁶ 735 F.3d at 742-43; see Jerre B. Swann, *Judge Richard Posner and Consumer Surveys*, 104 TMR 918 (2014). I have criticized Judge Posner’s observations as to trademark cases several times herein. As, however, Editor-in-Chief of *The Trademark Reporter*, it was my honor to republish a version of his article with William Landes, *Trademark Law: An Economic Perspective*, which I deem to be a major contribution to trademark law and required reading by anyone who would litigate trademark matters. See William M. Landes & Richard A. Posner, *The Economics of Trademark Law*, 78 TMR 267 (1988) (adapting William M. Landes & Richard A. Posner, *Trademark Law: An Economic Perspective*, 30 J. of L. & Econ. 265 (1987)).

⁹⁷ Jerre B. Swann, A “Reading” Test or a “Memory” Test: Which Survey Methodology Is Correct?, 95 TMR 876 (2005); Mike Rappeport, *Response to Survey Methodology Articles*, 96 TMR 769 (2006).

⁹⁸ Henry D. Ostberg, Ph.D, *Response to the Article Entitled “A ‘Reading’ Test or a ‘Memory’ Test: Which Survey Methodology Is Correct?”*, 95 TMR 1446 (2005).

⁹⁹ Cf. *Cumberland Packing Corp. v. Monsanto Co.*, 32 F. Supp. 2d 561, 578 (E.D.N.Y. 1999) (removing sweetener packet stimuli).

¹⁰⁰ It is my view that the “Ever-Ready” labels on the lamps and minibulb packets led to “batteries” responses to the “other products” question in that study.

¹⁰¹ Zeisel & Kaye, *supra* note 89, at 158-59; see *infra* note 139 and accompanying text.

Otherwise, the Eveready format—(i) particularly as now released from “top of mind” shackles and (ii) having moved beyond early “open-ended-questions-do-not-require-controls” thinking¹⁰²—is fully formed and well deserving of its gold standard status. Notwithstanding Judge Posner’s remarks in *Indianapolis Colts, Inc. v. Metropolitan Baltimore Football Club Ltd.* (intemperate in my view) that “[m]any experts are willing for a generous . . . fee to bend their science in the direction from which their fee is coming,”¹⁰³ the Eveready format is so comprehensible and the beneficiary of such a plethora of precedent that it is virtually tamper-proof. To repeat, it accurately measures whether a senior mark is sufficiently close to the surface of memory so as to be appreciably triggered by an allegedly similar junior use. Courts like open-ended questions and the Eveready format should be: (i) deployed wherever circumstances of accessibility, similarity, salience, familiarity, aided awareness and/or uniqueness permit; and (ii) as a memory probe to test accessibility in the *mind* in cases where the marks are not competitively proximate in the *market*.

In her concurrence to *Jack Daniel’s Properties, Inc. v. VIP Products LLC*,¹⁰⁴ Justice Sotomayor states that “in the context of parodies . . . , courts should treat the results of surveys with particular caution . . . [because s]urvey answers may reflect a mistaken belief among some survey respondents that all parodies require permission from the owner of the parodied mark”; and she quotes a “why do you say that” answer in the Eveready study executed in that case to illustrate her concern: “I’m sure the dog toy

¹⁰² See *supra* note 35.

¹⁰³ 34 F.3d 410, 415 (7th Cir. 1994). Survey experts do sometimes erroneously become advocates, mimicking their legal counterparts, but (with two exceptions) I personally have found them to be an intelligent and honorable lot. As for Judge Posner, I consider his contribution to the economics of trademark law to rank with Dr. Jacoby’s contributions to the cognitive underpinnings of likelihood of confusion surveys. I am not impressed, however, with Judge Posner’s understanding of survey science. In *Indianapolis Colts*, as a very minor (but telling) example, he agreed with a critique that “Baltimore Horses” as a control, “loaded the dice.” *Id.* at 415-16. Given, however, the *Buffalo Bills*, *Indianapolis Colts* and *Birmingham Stallions*, Dr. Jacoby did not have a wealth of alternatives. I do believe that Dr. Jacoby distorted the stimuli in *Kargo Global, Inc. v. Advance Magazine Publishers, Inc.*, No. 06 Civ. 550(JFK), 2007 WL 2258688 (S.D.N.Y. Aug. 6, 2007) (discussed in *Cognitively Updated*, *supra* note 6, at 747-48), but in *Indianapolis Colts*, he was working with a limited palette and “Horses” was better than donkeys, mules or jackasses—other four-legged equines offered in thesauri—and, particularly given the horse head in the *Indianapolis Colts* logo, better than Posner’s suggestion of the “Baltimore Leopards.” *Indianapolis Colts*, 34 F.3d at 416. As I note *infra*, the selection of a control is frequently a *sui generis* exercise and often involves more art than science. I am confident that “Baltimore Horses” did not filter all the noise from Dr. Jacoby’s *Indianapolis Colts* survey, but I am likewise confident that Jacoby was not there “bend[ing] his] science in the direction from which [his] fee [was] coming” as he may have done in *Kargo Global* (where to his credit his control filtered all the noise that his distorted test had generated).

¹⁰⁴ 143 S. Ct. 1578 (2023).

company that made this toy had to get [Jack Daniel's] permission . . . to essentially copy the[ir] product in dog toy form.”¹⁰⁵

I read the *Jack Daniel's* decision shortly after it was announced but did not initially regard Justice Sotomayor's concerns as significant. I understood her effectively to be saying that an Eveready study should be considered in the context of an overall confusion factors analysis in assessing likelihood of confusion, which I have always understood to be an Eveready imperative. A senior company response to a “who makes or puts out” question is not evidence of actual confusion but may only support an inference as to likelihood of confusion.¹⁰⁶ The Eveready format itself is grounded in the likelihood of confusion factors of: (i) (mental) *proximity* and (ii) *similarity* that drive any likelihood of confusion assessment (whether Eveready or Squirt). To iterate, as I write above:

an Eveready study is a reliable measure of ***real-world accessibility*** [availability for a comparison in the mind] ***and similarity***, *i.e.*, of whether the senior use is sufficiently close to the surface of memory, and the junior use is sufficiently similar to the senior use so that . . . the former will likely be activated (recalled) by a monadic exposure to the latter.

Because it is confusion factor driven, I have never encountered difficulty in conducting a likelihood-of-confusion factors analysis in the context of a positive Eveready result.

Upon, however, a further reading of Justice Sotomayor's concurring opinion in the peer review process for this article, I additionally believe that she was in error in her parody analysis. “A parody must convey two simultaneous — and contradictory — messages: that it is the original, but also that it is not the original and is instead a parody.”¹⁰⁷

The “why answer” that Justice Sotomayor quotes reflects that the respondent had conveyed that it was the original; by “copy[ing]” Jack Daniel's mark as its own mark, VIP Products clearly accessed the Jack Daniel's schema. VIP failed, however, to communicate that the dog toy was not the original (or authorized by Jack Daniel's) and was not only “a poor parody but also vulnerable under trademark law, since the customer [would] be confused.”¹⁰⁸

¹⁰⁵ *Id.* at 1593 (Sotomayor, J., concurring).

¹⁰⁶ *See, e.g.*, *Union Carbide Corp. v. Ever-Ready, Inc.*, 531 F.2d 366 (7th Cir. 1976). In that case, the Seventh Circuit discusses survey results, as noted above, in the context of an incontestable mark with “overwhelming” evidence of secondary meaning, with which consumers were “thoroughly acquainted,” where there was evidence of actual confusion, and the marks were legally and aurally identical. *Id.* at 380-85.

¹⁰⁷ *Cliffs Notes, Inc. v. Bantam Doubleday Dell Publ'g Grp., Inc.*, 886 F.2d 490, 494 (2d Cir. 1989).

¹⁰⁸ *Id.*

Alternatively, Justice Sotomayor would use why answers potentially to establish causation: i.e., that confusion was a result of mistaken beliefs as to a parodist's need for permission. As I note above, "a *consistent pattern*, across multiple questionnaires and thousands of respondents, of 35% or more of respondents stating that Adidas was the maker of a two/four stripe athleisure shoe, with many giving 'stripes' as the reason for their answer, reflected that respondents were accessing the adidas three stripe brand schema in their memory, and that pattern matching was the most plausible explanation for their response that Adidas made or put out a two/four stripe athleisure shoe"; i.e., why answers (under rarely existing circumstances) may inform causation.

As I state below, however, in my discussion as to the survey conducted in *SquirtCo*, using why answers to prove causation is impermissible:

. . . [F]ifty-three of four hundred seventy-six respondents (11 percent) gave "different compan[y]" responses, but answered the "what makes you say that" question in a manner that the court interpreted as manifesting confusion and were added to the sixty-one (13 percent) "same company" respondents. In early Squirt format cases, same company responses were sometimes counted as confusion only if a trademark relevant reason was given to a "why do you say that" question, but with the advent of scientific methodologies, it has been appreciated that only by utilizing controls can a survey prove causation. The court thus erred in [giving] "what makes you say that" answers [a causal impact by switching] "different compan[y]" answers to the "same company" column. . . .

I am presently uncertain how to design a control to filter out answers prompted by a perceived need to get permission (mere "noise" in Justice Sotomayor's view) from responses prompted by the mental accessibility (*proximity*) and *similarity* factors in a likelihood of confusion analysis, but again it is a parodist's obligation to negate sponsorship, not a brand owner's obligation to show that (i) similarity of information (ii) in a respondent's memory is alone, or the predominant, cause of confusion. Justice Sotomayor has thus effectively placed the burden of proof shoe on the wrong foot.

A. Subsequent Applications of the Eveready Format

Throughout the 20th century, Welter's top-of-mind requirement prevailed as a predicate for Eveready usage. In *Simon & Schuster, Inc. v. Dove Audio, Inc.*,¹⁰⁹ (THE BOOK OF VIRTUES publications v. THE CHILDREN'S BOOK OF VIRTUES publication), for example, defendant countered plaintiff's Squirt study reflecting 36%

¹⁰⁹ 970 F. Supp. 279 (S.D.N.Y. 1997).

confusion with an Eveready-type (unaided) study reflecting 6.6% confusion, but the court intoned that defendant's survey could have significantly underestimated confusion because the Eveready format required top-of-mind awareness which plaintiff's book did not possess.¹¹⁰

The benefit of the Eveready format was thus relegated to mega-brands like the MCDONALD'S mark. In *McDonald's Corp. v. McBagel's, Inc.*,¹¹¹ for example, 24.8% of respondents answered that McDonald's sponsored or promoted a MCBAGEL'S restaurant.¹¹² In *Quality Inns International, Inc. v. McDonald's Corp.*,¹¹³ 30% of respondents answered that McDonald's owned or operated MCSLEEP INN (58% of whom gave "Mc" in the respective names as their major reason for so responding).¹¹⁴ In *McDonald's Corp. v. Druck & Gerner, DDS., P.C.*,¹¹⁵ 30% of respondents answered McDonald's to the question "[w]ho or what company do you believe owns or operates McDental?"¹¹⁶

Druck & Gerner was included in the text of the book that I used to teach a Trademark course at the University of Georgia and, after the session on likelihood of confusion factors (stressing the commercial strength factor), I annually asked (softly so as suggest that an affirmative response could reflect *not* ignorance, but a student's understanding as to the dominance and "reach" of the MCDONALD'S brand) for a show of hands by anyone who might think that McDonald's operated a dental practice. In seven years, no hand was raised. Something other than a traditional likelihood of confusion was afoot.

As I stated in "Swann's Response to Diamond Re Dilution Surveys" in the Second Edition, "[l]ikely confusion and likely dilution are [sometimes] kindred concepts":

[T]he defendant in *Federal Express Corp. v. Federal Express, Inc.*, admitted thinking that "Federal Expresso" would be an easy name to remember [for a coffee shop] because of its similarity to "Federal Express," which she knew to be a well-known registered trademark. Similarly, the defendants in *McDonald's Corp. v. Druck and Gerner, DDS*, admitted that

¹¹⁰ *Id.* at 298-99.

¹¹¹ 649 F. Supp. 1268 (S.D.N.Y. 1986).

¹¹² *Id.* at 1277.

¹¹³ 695 F. Supp. 198 (D. Md. 1988) (an Eveready as to MCTRAVEL ("Mc" + generic service category) produced confusion; MCTAVISH ("Mc" + a proper name) did not; the reach of the "Mc" prefix is thus limited to a commercial context).

¹¹⁴ *Id.* at 208.

¹¹⁵ 814 F. Supp. 1127 (N.D.N.Y. 1993).

¹¹⁶ *Id.* at 1134 & n.8. In none of the McDonald's cases was there a control for brand dominance. I am certain that many of the 24.8% of respondents who answered that McDonald's sponsored or promoted McBagel's did so reflexively—in the 1980s, McDonald's dwarfed the fast-food industry.

“McDental” had a quality of “retentiveness” and was “memorable.”¹¹⁷

Schechter [the father of dilution] would have characterized the injunctions in *Federal Express* and *McDental* as instances where courts were loath to permit defendants to “get the benefit of complainant’s reputation,” which he saw as a “salutary . . . recognition of the actual nature and function of the trade-mark under [1927] conditions.”¹¹⁸

A variant of free riding on the retentiveness and memorability of the “Mc” prefix in a commercial context rather than traditional confusion would have thus been Schechter’s concern and is the true rationale for the MCDENTAL mark result—even a dentistry could tap into the recall prowess of a brand like MCDONALD’S.¹¹⁹

Skipping to *Amstar Corp. v. Domino’s Pizza, Inc.*,¹²⁰ plaintiff surveyed female heads of household primarily responsible for making food purchases and asked what other products were made by the company that put out DOMINO’S pizza. Seventy-one per cent answered sugar.¹²¹ The Fifth Circuit rejected the survey for universe considerations: respondents in a likelihood of confusion study should be prospective purchasers of defendant’s, not plaintiff’s, products.¹²²

Had I known what I know today, I could also have argued that plaintiff inflated “sugar” responses by failing to control for brand dominance of DOMINO for sugar among heads of household responsible for making food purchases. Many grocery shoppers, I am certain, reflexively answered “sugar” when asked about other products made by the company that put out a container boldly marked DOMINO’S. The ghost of the failure in Eveready to control for brand dominance haunted likelihood of confusion studies for decades.

¹¹⁷ In *Quality Inns Int’l, Inc. v. McDonald’s Corp.*, the declaratory judgement plaintiff characterized the “McSleep” brand in an early brochure as “instantly recognizable.” 695 F. Supp. 198, 204 (D. Md. 1988).

¹¹⁸ Jerre B. Swann, *Swann’s Response to Diamond Re Dilution surveys*, in Second Edition, *supra* note 7, at 176.

¹¹⁹ In one of the premier dilution articles that I have read, Robert G. Bone abjures free riding as a dilution rationale. Robert G. Bone, *Hunting Goodwill: A History of the Concept of Goodwill in Trademark Law*, 86 Boston Univ. L. Rev. 547 (2005). As a member of the Special Committee with respect to the Trademark Dilution Revision Act of 2006 and in my article “Dilution Redefined for the Year 2000,” 90 TMR 823 (2000), I advocated free riding as a dilution factor consistent with Schechter’s comment that some courts were loath to permit defendants to “get the benefit of complainant’s reputation.”

¹²⁰ 615 F.2d 252 (5th Cir. 1980).

¹²¹ *Id.* at 263-64.

¹²² *Id.* at 264; accord William G. Barber & Giulio E. Yaquinto, *The Universe*, in Second Edition, *supra* note 7, at 32-34.

In perhaps the most consequential Eveready of the 20th century, plaintiff in *Qualitex Co. v. Jacobsen Products Co.*,¹²³ showed respondents in the dry-cleaning industry green-gold ironing pads and 39% responded that Jacobsen was the manufacturer.¹²⁴ The Ninth Circuit held that color alone was not protectable;¹²⁵ the Supreme Court reversed and noted, as did the Seventh Circuit in *Union Carbide*, that *the Eveready survey result was evidence of secondary meaning*.¹²⁶

In the 21st century, the well-established, “widely accepted” Eveready format¹²⁷ has clearly emerged as the preferred methodology for testing “commercially strong” marks,¹²⁸ particularly where the marks are not competitively proximate.¹²⁹ Eveready surveys have been frequently relied on to support an inference as to likelihood of confusion, and a well-done Eveready extends not only to the junior marks tested, but to any others that share “common and prominent features.”¹³⁰ A defendant cannot financially exhaust a prominent senior by insisting that each later infringing edition or minor variation be separately tested.

Of equal importance was the beginning erosion early in the 21st century of the top-of-mind standard¹³¹: as noted above, accessibility now ranges from top-of-mind to aided awareness. A mark that is *not in memory cannot be accessed by an Eveready probe format*, but a multiple of factors—salience, uniqueness, aided awareness, familiarity—inform as to how far below the surface of memory an Eveready may penetrate; and, as to marks that are not

¹²³ 514 U.S. 159 (1995).

¹²⁴ *Qualitex Co. v. Jacobsen Prods. Co.*, 13 F.3d 1297, 1305 (9th Cir. 1994).

¹²⁵ *Id.* at 1301-02.

¹²⁶ 514 U.S. at 163-64.

¹²⁷ Robert Kirkman, LLC v. Toking Dead, 2020 WL 4365811, at *10 (T.T.A.B. July 22, 2020); Combe Inc. v. Dr. August Wolff GmbH & Co. KG Arzneimittel, 382 F. Supp. 3d 429, 462 (E.D. Va. 2019); Coty, Inc. v. Excell Brands, LLC, 277 F. Supp. 3d 425, 451 (S.D.N.Y. 2017).

¹²⁸ See *Gilead Scis., Inc. v. Gilead Cap. LP.*, 2021 WL 1592674, at *32-34 (T.T.A.B. Apr. 20, 2021); *PepsiCo, Inc. v. Pirincci*, 2012 WL 2930650, at *8 (T.T.A.B. June 25, 2012).

¹²⁹ See *Icleen Entwicklungs-Und Vertiebsanstalt Für Umweltprodukte v. Blueair AB*, No. CV 21-2236 DSF (ADS), 2021 WL 6104397, at *8 (C.D. Cal. Oct. 4, 2021); see also *Denimafia Inc. v. New Balance Athletic Shoe, Inc.*, No. 12 Civ. 4112(AJP), 2014 WL 814532, at *20 (S.D.N.Y. Mar. 3, 2014) (which contains an excellent discussion of the impact of brand strength as well as the failure of a “sequential array survey” to replicate the marketplace in the “absence of evidence that . . . products would be encountered in close proximity”).

¹³⁰ *Coty, Inc.*, 277 F. Supp. 3d at 452; *Adidas Am., Inc. v. Payless Shoesource, Inc.*, 529 F. Supp. 2d 1215, 1227 (D. Or. 2007) (“Where actually surveyed products and subsequently accused products share common and prominent features, a trademark infringement plaintiff need not create new likelihood of confusion surveys for each newly accused product.”).

¹³¹ See *Wizkids/NECA, LLC v. TIII Ventures, LLC*, No. 17-CV-2400 (RA), 2019 WL 1454666, at *12 (S.D.N.Y. Mar. 31, 2019).

competitively proximate, proving that a mark is not in the mind at all is outcome determinative.¹³²

The most satisfying Eveready survey that I commissioned was in *H-D Michigan, Inc. v. Top Quality Service, Inc.*¹³³ Defendant conducted Caribbean cruises under the banner of HOGS ON THE HIGH SEAS which generated actual confusion among motorcycle owners (e.g., defendant was frequently asked “do I need to own a Harley to go on a cruise?”). I obviously could not use the Squirt format because the marks involved were not competitively proximate, but few marks have the *reach* of HARLEY DAVIDSON and its panoply of brands, and the Eveready format is ideal for testing marks with reach. I employed the same expert who had done the surveys in *Facebook* and in the two/four stripe *Adidas* cases and his survey: (i) “suggest[ed] that twenty-seven percent of motorcycle riders . . . believe[d] that Harley promote[d] or approve[d] of [defendant’s] cruises [and] [ii] that much of this confusion stem[med] from [defendant’s] use of the word ‘hog.’”¹³⁴ Because, however, “hog” had been declared to be generic for large *motorcycles* in *Harley-Davidson, Inc. v. Grottanelli*,¹³⁵ the district court refused to enjoin defendant’s conduct.¹³⁶

The Seventh Circuit reversed, noting that “much of this confusion stem[med] from . . . the word ‘hog’[s] . . . association with the Harley Owners Group” [H.O.G., a 700,000-member social club of Harley owners].¹³⁷ The survey results were self-validating in that respondents were accessing (i) Harley Owners Group or H.O.G. when asked who promoted a HOGS ON THE HIGH SEAS Caribbean tour; (ii) the Second Circuit had not ruled that H.O.G. or “hog” was generic for *motorcyclists*, and (iii) the District Court’s reliance on *Grottanelli* to deny relief was thus misplaced. The ability of the Eveready format to probe memory as to the full panoply of brands associated with Harley-Davidson and the word “hog” was crucial.

B. Variants of the Eveready Format

Eveready variants have ranged from the bizarre to the sublime. In the bizarre category was *Arche, Inc. v. Azaleia, U.S.A., Inc.*:

Plaintiff’s counsel designed a questionnaire and sent one of their employees . . . , a part-time typist, drama student and actress, into Washington Square Park, which is located within blocks of one of plaintiff’s retail stores. Over a two[-

¹³² See *supra* note 87 and accompanying text.

¹³³ 496 F.3d 755 (7th Cir. 2007).

¹³⁴ *Id.* at 758.

¹³⁵ 164 F.3d 806 (2d Cir. 1999).

¹³⁶ 496 F.3d at 759.

¹³⁷ *H-D Michigan*, 496 F.3d at 758.

l day period, she approached a number of people who, she said, looked as if they could afford plaintiff's shoes, which sell at prices considerably higher than defendants. [The employee] who was shod in defendants' shoes, asked the well-to-do passersby whether they could identify the shoes she was wearing. Although she had a questionnaire designed by counsel, she did not necessarily ask the questions as written. All told, she interviewed forty-six people. That she hoped to find people who identified defendants' product as Arche shoes is evident both from the selection of higher income subjects and from the questionnaire, which instructed her to try to enlist persons giving the desired "right" answer as witnesses.¹³⁸

In the sublime category was the implementation by Dr. Zeisel of the format that he outlined in *Proving it with Figures*¹³⁹ in connection with a dispute as to the PAM and PAN-LITE cooking spray brands:

Dr. Zeisel supervised two surveys in which only primary shoppers over eighteen years of age were questioned. In the first survey, participants were shown photographs of six retail grocery items commonly sold in supermarkets. These items included Pan-Lite but not PAM. After seeing the photographs, participants were asked to recall the brand names they had seen. Forty-five percent of those recalling a brand name for the cooking spray identified Pan-Lite as PAM.

The second survey was performed as a control because PAM is the product leader in the market and might be identified by consumers for that reason. The control survey showed participants the same six grocery items, but substituted Mazola or Wesson-Lite (also direct competitors of PAM) for Pan-Lite. Seven percent of the participants in the second survey identified the cooking spray as PAM. Dr. Zeisel then subtracted the seven percent figure from the forty-five

¹³⁸ 882 F. Supp. 334, 335 (S.D.N.Y. 1995).

¹³⁹ *Supra* note 89. I was fortunate that my firm retained Dr. Zeisel to advise it with respect to surveys in the *Frito Lay/Jays Foods* matter in which Jays unsuccessfully contended that FRITOS was generic for corn chips, and I am still in awe of his expertise as to consumer research which was a major spark for my interest in genericism studies as well as in likelihood of confusion studies. See Jerre B. Swann, *The Economic Approach to Genericism: A Reply to Folsom and Teply*, 70 TMR 243 (1980); Jerre B. Swann, *Genericism Rationalized*, 89 TMR 639 (1999). As one downside, Dr. Zeisel insisted on probability studies and lowly associates (as then was I) were assigned to go to designated residences to conduct pilot interviews in the Atlanta area. I spent one summer month fending off snarling dogs, but as the ultimate result of many pilot studies, Frito Lay adopted what was to become the *Teflon* format for assessing genericness. The case was settled before a survey was required so that my efforts along with those of three or four others went for naught, but it was a summer that I will never forget.

percent figure to control for the likelihood of misidentification that could reasonably be considered a result of PAM's status as market leader. The result was a net misidentification of thirty-eight percent attributable to the similarities between Pan-Lite and PAM.¹⁴⁰

The district court erroneously rejected the Zeisel study for failing to show PAM and PAN-LITE as found in the marketplace “shelved together . . . or amongst [other] cooking spray[s]”¹⁴¹ (as they might be presented in a Squirt format study). As an *Eveready* variant, however, showing respondents only a contextual depiction of defendant's mark to trigger (or not) a pattern matching search of memory, is universally approved and the Seventh Circuit thus deemed the rejection “premature” at the summary judgment stage. In my view, the Zeisel methodology is an excellent replication of a consumer's experience “shopping for frequently bought household goods and confronted with a number of different brands [making] quick decisions based on the ‘gestalt’ of the product.”¹⁴²

Also in the sublime category was the *aided* Eveready¹⁴³ executed by the defendant in *Sazerac Co. v. Fetzer Vineyards, Inc.*¹⁴⁴ In that case, plaintiff alleged that its (admittedly) weak trade dress of: (i) its BUFFALO TRACE bourbon in (ii) squat, tannish bottles with a green capsule, bearing its (iii) Buffalo Logo; was infringed by defendant's: (i) 1000 STORIES “Bourbon Barrel Aged” zinfandel in (ii) tall, black bottles with a red capsule, bearing (iii) a buffalo (nickel) image. Of different alcoholic content and for use on different occasions,¹⁴⁵ the products were both sold in liquor and grocery stores, but not on the same shelf or aisle; they were *not competitively proximate*.

Plaintiff's expert nonetheless utilized a Squirt two-room format: in the first room, he exposed respondents to BUFFALO TRACE

¹⁴⁰ AHP Subsidiary Holding Co. v. Stewart Hale Co., 1 F.3d 611, 613-14 (7th Cir. 1992) (footnote omitted).

¹⁴¹ *Id.* at 615.

¹⁴² Zeisel & Kaye, *supra* note 89, at 158.

¹⁴³ An aided Eveready is essentially a Squirt with Eveready open-ended questions. See *Cognitively Updated*, *supra* note 6, at 730 n.13 (“[W]here the two marks are typically seen in the marketplace side-by-side [or in close proximity, and] the researcher desires to use the open-ended Eveready questions[,] . . . [the] researcher might present respondents with a store display that includes . . . the two at issue as well as representative others[:] after allowing respondents to evaluate the products on display, the allegedly infringing product . . . is placed in front of the respondent; the sequence then follows the Eveready question protocol (while leaving the other products in sight). Such a methodology . . . maintains the marketplace reality without asking potentially leading questions.”); Itamar Simonson & Ran Kivetz, *Demand Effects in Likelihood of Confusion Surveys: The Importance of Marketplace Conditions*, in First Edition, *supra* note 7, at 243-59.

¹⁴⁴ 265 F. Supp. 3d 1013 (N.D. Cal. 2017).

¹⁴⁵ *Id.* at 1035.

bourbon for (a forced and unrealistic) five seconds and in the second room, he showed “an array featuring zinfandels including 1000 Stories . . . creat[ing] a situation that’s ‘not terribly likely’ to happen in the [liquor and grocery store] marketplace.”¹⁴⁶ “The methodology [did] not remotely approximate the marketplace.”¹⁴⁷ In my view, the 46% confusion that the survey produced, as to bottles that in the district court’s view were not similar, redounded to plaintiff’s detriment.¹⁴⁸

Defendant’s expert, for the purpose of a comparison, first exposed respondents to an *aided* stimulus including plaintiff’s labeled bourbon bottle, a 1000 STORIES bottle, and a control bottle among thirteen “bottles that a shopper might see at a grocery or liquor store,” before asking who put out a sequentially displayed 1000 STORIES bottle.¹⁴⁹ Finding that the defendant’s survey stimulus more closely approximated conditions of sale—and the *unlikelihood* (in my view), given the cluttered marketplace, that not-directly-competing or adjacent brands would simultaneously be in cognitive workspace for a comparison—the court adopted the finding of defendant’s expert of “absolutely no confusion.”¹⁵⁰

Competitive proximity is required to justify aiding and sequential presentations in the Squirt format.¹⁵¹ Bourbon and zinfandel were in the same stores, but they were not competitively proximate. Nowhere in the real world did there likely exist a sequential line-up of BUFFALO TRACE bourbon and zinfandels; artificial proximity inflated the likelihood of confusion percentages and demand effects abounded.¹⁵²

III. THE HISTORY OF THE SQUIRT FORMAT

Recall from memory, to iterate, can be a difficult cognitive exercise¹⁵³ and many marks either are: (i) not sufficiently **accessible in memory** or are (ii) not “infringed” by a sufficiently

¹⁴⁶ *Id.* at 1026.

¹⁴⁷ *Id.*

¹⁴⁸ Although I disagree with certain remarks of Dr. Daryl Lim as to the cost (burden) of surveys in the internet age, it has been my experience and I fully concur that “judges [often] . . . reject surveys based on whether the results agree with their subjective impressions.” Daryl Lim, *Trademark Confusion Revealed: An Empirical Analysis*, 71 *Am. Univ. L. Rev.* 1285, 1313 (2022). In my experience, moreover, a party materially damages a court’s perception of the merits of its case by offering, as did Sazerac, clearly inflated survey results as to likelihood of confusion. A poor, slanted survey may be far worse than no survey at all.

¹⁴⁹ 265 F. Supp. 3d at 1027-28.

¹⁵⁰ *Id.* at 1028.

¹⁵¹ See *Ever-Constant Eveready*, *supra* note 6, at 675-76.

¹⁵² For demand effects, see Simonson & Kivetz, *supra* note 143.

¹⁵³ See Jacob Jacoby, *Are Closed-Ended Questions Leading Questions?*, in *First Edition*, *supra* note 7, at 269-72.

similar junior mark so as to trigger “aided” awareness/**recall** of the senior mark. Where, however, marks are competitively proximate in the market so as to make an aided presentation of both marks to respondents reflective of **marketplace reality**, the marks may be so similar that respondents (and consumers) react that they come from the “same company.” Thus, as to competitively proximate marks, the (i) *aided* Squirt format, with (ii) a *closed-ended* question, may replicate a real-world comparison (iii) **in the marketplace** when respondents are asked (iv) whether contextually presented *similar* marks come from the same or a different company.

Aiding in the Squirt format, to replicate the real-world marketplace, inserts (in effect) an otherwise inaccessible mark into the mind for a similarity/fit (pattern matching) assessment as occurs in the Eveready format with respect to marks that are accessible in memory. Aiding thus reflects a consumer’s experience of: (i) seeing a senior mark; (ii) storing it in fleeting memory;¹⁵⁴ and encountering a similar junior mark for comparison purposes before memory of the senior mark has degraded. Because forgetting begins virtually immediately,¹⁵⁵ “with the most rapid decrement occurring earlier rather than later” (November 1, 2014, email from Dr. Jacob Jacoby), products must effectively be *competitively proximate to support pattern matching with respect to “aided” presentations of marks not otherwise readily accessible in memory*. “The controlling consideration is whether consumers are likely to see an allegedly infringing use sufficiently close in time or space to a senior use so that both are sufficiently still in (non-degraded) memory so as to permit their comparison,”¹⁵⁶ and I agree with Dr. Jacoby that courts have adopted “competitive proximity” for that purpose.¹⁵⁷

Effectively, to iterate, a lesser known mark (less or non-accessible in memory) may *reach* to a similar junior use where the marks are so proximate that one of the marks at issue is likely still

¹⁵⁴ For likelihood of confusion purposes, it matters not, of course, that a consumer might actually first see and store in transient memory the junior mark and then encounter the senior mark before memory of the junior decays.

¹⁵⁵ Alan Baddeley, *Hum. Memory: Theory & Prac.* 169 (1990).

¹⁵⁶ *Cognitively Updated*, *supra* note 6, at 744; *see* *Moroccanoil, Inc. v. Zotos Int’l, Inc.*, 230 F. Supp. 3d 1161, 1175 (C.D. Cal. 2017) (The Squirt design is “most appropriate where a product with a weak mark is sold in close proximity to the alleged infringer in the marketplace.” . . . Here, consumers may encounter both marks in close proximity on the Internet on the same websites Nonetheless, because the products sold on these websites are non-authorized and it is unclear if a significant number of consumers would encounter the marks in close proximity, the Court gives the survey less weight.” (citations omitted)).

¹⁵⁷ Jacob Jacoby, *Trademark Surveys* § 3.12.1 (2013). Geographic proximity can be a factor in service industries where “competing” entities can exist “across great distances.” *See* *Isle of Capri Casinos, Inc. v. Flynt*, No. 2:16-cv-06148-CAS (MRWx), 2016 WL 6495380, at *6-7 (C.D. Cal. Nov. 1, 2016) (plaintiff’s Squirt survey rejected for failure to replicate marketplace conditions as to plaintiff’s casino located in Black Hawk, Colorado, and defendant’s casino in Los Angeles).

to be in a consumer's cognitive workspace for comparison when the other is encountered.¹⁵⁸ In *SquirtCo*,¹⁵⁹ for example, replicating the (potential) competitive proximity (in Chicago) of the products there at issue, survey respondents first heard radio ads for **both** SQUIRT and QUIRST soft drinks;¹⁶⁰ and were then asked (i) the **closed-ended** question, "Do you think SQUIRT and QUIRST are put out by the same company or by different companies?"; followed by (ii) the question, "What makes you think that?"¹⁶¹

Squirt was "coined" as a trademark for soft drinks (specifically a carbonated grapefruit drink) in 1937, "apparently as a reference to the tendency of grapefruit to squirt when eaten with a spoon," and was registered by SquirtCo in 1939.¹⁶² By 1980, it was a strong, distinctive mark¹⁶³ that could (and likely should) have been tested using the Eveready format had *Union Carbide* been decided at the time that SquirtCo's dispute with Seven-Up arose.

Quirst was "coined" in 1977, as (Seven-Up claimed) a "neologism" made up of "**quen**ch **thirst**."¹⁶⁴ As evidence of Seven-Up's intent, SquirtCo insisted that Quirst was merely an anagram of Squirt.¹⁶⁵ The district court, siding with Squirt's view, found that there was "similarity in sound between . . . SQUIRT and QUIRST" and "extensive similarity" between the marks.¹⁶⁶

Given (i) the commercial strength of the Squirt mark; (ii) the similarity of the word marks; (iii) the similarity and "competitive proximity" of the products; and (iv) the low degree of care likely to be exercised by purchasers of soft drinks, the district court correctly observed that there were "signs of a likelihood of confusion."¹⁶⁷ The district court and the Eighth Circuit also relied (erroneously) on SquirtCo's "surveys" in Chicago and Phoenix¹⁶⁸ which were seriously flawed in multiple respects:

1. The juxtaposition of only two marks in conjunction with a closed-ended question, as noted above, has been repeatedly and properly determined to be leading;¹⁶⁹ and immediate juxtaposition has been substantially replaced by a sequential

¹⁵⁸ *Ever-Constant Eveready*, *supra* note 6, at 673.

¹⁵⁹ 628 F.2d 1086 (8th Cir. 1980).

¹⁶⁰ *Id.* at 1089 n.4. In Phoenix, respondents only heard *abstractly* the SQUIRT and QUIRST names. *Id.* In neither market were respondents shown the respective product bottles.

¹⁶¹ *Id.*

¹⁶² *Id.* at 1088.

¹⁶³ *Id.* at 1089-90 ("A strong and distinctive trademark is entitled to greater protection than a weak or commonplace one.").

¹⁶⁴ *Id.* at 1088 & n.1 (emphasis added).

¹⁶⁵ *Id.* at n.1.

¹⁶⁶ 628 F.2d at 1089-90.

¹⁶⁷ *Id.* at 1089-90.

¹⁶⁸ *Id.*

¹⁶⁹ *Supra* note 36 and accompanying text.

“two room” line-up format¹⁷⁰ that in McCarthy’s view “is an attempt to replicate the marketplace process of advertising exposure to a brand or trade dress, followed by being confronted in the market with both similar and differing brands or trade dresses.”¹⁷¹ The use of only two stimuli has been universally replaced by the use of four or more stimuli in a “second room” or in an array (one of which may serve as an in-treatment control) to “remove[] the spotlight from the products of the plaintiff and defendant, help[] avoid making obvious what the survey is about, and make[] the survey more realistic and less leading.”¹⁷²

2. As Seven-Up observed, “soft drinks [were] selected in the marketplace by visual recognition.”¹⁷³ SquirtCo’s surveys, however, “were based on sound alone, with no visual component.”¹⁷⁴ As McCarthy notes, “[t]he closer the survey methods [and stimuli] mirror the situation in which the ordinary person would encounter the trademark, the greater the evidentiary weight of the survey results.”¹⁷⁵ Context is critical.¹⁷⁶ As I note in Second Edition, *supra* note 7, at 78, while few surveys are rejected for choice of format, “[m]any are rejected for the abstract presentation of stimuli.” SquirtCo’s use of the word marks as stimuli, rather than the

¹⁷⁰ See, e.g., *Storck USA, L.P. v. Farley Candy Co.*, 797 F. Supp. 1399, 1408 (N.D. Ill. 1992).

¹⁷¹ McCarthy, *supra* note 23, § 32:177.

¹⁷² *Tokidoki, LLC v. Fortune Dynamic, Inc.*, No. CV 07-1923 DSF (PJWx), 2009 WL 2366439, at *8 (C.D. Cal. July 28, 2009).

¹⁷³ *SquirtCo v. Seven-Up Co.*, 628 F.2d 1086, 1089 n.4 (8th Cir. 1980). When I last regularly purchased COKE, PEPSI, ORANGE CRUSH and RC soft drinks (always with a MOON PIE marshmallow sandwich) in the early 1960s, I differentiated the products by their bottles, and the green clunky COKE bottle was perhaps the strongest mark in the cola category (depictions of which notably survive as a mark even in the aluminum can era).

¹⁷⁴ *Id.*; *accord M.D. On-Line, Inc. v. WebMD Corp.*, No. 05-CV-4081 (WJM), 2005 WL 2469668, at *7 (D.N.J. Oct. 6, 2005) (“Here, plaintiff’s survey . . . was done over the telephone, participants were not shown the marks . . . Thus, any alleged association the participants found between the marks was based on sound alone, and does not address how the marks might be viewed in a commercial setting.”); *Beverage Mktg., USA, Inc. v. S. Beach Beverage Corp.*, No. 97CIV.4137(LMM), 2000 WL 1708214, at *2 (S.D.N.Y. Nov. 15, 2000) (“Comparing the bottles with the labels properly on them . . . the Court holds that a reasonable jury . . . could not find a likelihood of confusion.”).

¹⁷⁵ McCarthy, *supra* note 23, § 32:163.

¹⁷⁶ At one point, I considered listing in a footnote all the cases discussing the criticality of context, but the list rapidly ran for pages, and I decided to refer only to a footnote in *Cognitively Updated*, *supra* note 6, at 746 n.76, which discussed several of the more cogent rulings: *inter alia*, *Water Pik, Inc. v. Med-Systems, Inc.*, 726 F.3d 1136, 1145-47 (10th Cir. 2013); *J.T. Colby & Co. v. Apple Inc.*, No. 11 CIV. 4060 DLC, 2013 WL 1903883, at *20-22 (S.D.N.Y. May 8, 2013), *aff’d*, 586 F. App’x 8 (2d Cir. 2014); *Steak Umm Co. v. Steak ‘Em Up, Inc.*, No. 09-2857, 2011 WL 3679155, at *6 (E.D. Pa. Aug. 23, 2011); *Pilot Corp. of Am. v. Fisher Price, Inc.*, 344 F. Supp. 2d 349, 358 (D. Conn. 2004); *Juicy Couture, Inc. v. L’Oreal USA, Inc.*, No. 04 Civ.7203(DLC), 2006 WL 1012939, at *23-27 (S.D.N.Y. Apr. 19, 2006).

labelled bottles which ordinary consumers would encounter in the marketplace, alone would have mandated (in my view) the rejection of its survey efforts.

3. The survey results were not presented by a survey expert and were improperly interpreted by the district court. In Phoenix, for example, fifty-three of four hundred seventy-six respondents (11 percent) gave “different compan[y]” responses, but answered the “what makes you say that” question in a manner that the court interpreted as manifesting confusion and were added to the sixty-one (13 percent) “same company” respondents.¹⁷⁷ In early Squirt format cases, same company responses were sometimes counted as confusion only if a trademark relevant reason was given to a “why do you say that” question,¹⁷⁸ but with the advent of scientific methodologies, it has been appreciated that only by utilizing controls can a survey prove causation. The court thus erred in using “what makes you say that” answers to switch “different compan[y]” answers to the “same company” column, and the error was material, given the Eight Circuit’s affirmance that approximately 25% same company responses (11%+ 13% in Phoenix) “was sufficient to support an inference of likelihood of confusion.”¹⁷⁹
4. “Seven-Up’s principal [(and wholly legitimate)] attack on the survey was that the first question was dichotomous and encouraged guessing by *suggesting* that it was improper to answer ‘I don’t know.’”¹⁸⁰ Today: (i) controls are universally used to filter guessing; (ii) there are admonitions not to guess; and (iii) there are express “don’t know” options.¹⁸¹

If I repeated as to the Squirt format, as I observed with respect to the Eveready format, that its origins were “clouded,” I would be engaging in euphemism. Plaintiff’s surveys that were relied on in *SquirtCo* were among the most flawed that I encountered in the thousands of cases and summaries¹⁸² that I have reviewed for this historical account.

¹⁷⁷ 623 F.2d at 1089 n.4.

¹⁷⁸ See, e.g., *Cognitively Updated*, *supra* note 6, at 742 n.60 & 749 n.81; cf. *Sally Beauty Co. v. Beautycos, Inc.*, 304 F.3d 964, 979-80 (10th Cir. 2002) (same company responses identifying “package/appearance” as the reason why provide “strong evidence of actual confusion”).

¹⁷⁹ *Id.* at 1089 n.4, 1091.

¹⁸⁰ *Id.* at 1089 n.4.

¹⁸¹ David T. Neal, *Psychological Considerations in Designing Trademark and False Advertising Survey Questionnaires*, in *Second Edition, supra* note 7, at 227; Jacob Jacoby, *Trademark Surveys* § 8.32 (2013).

¹⁸² Without access to the summaries in Welter’s *Trademark Surveys* and in “Intellectual Property Surveys Annual Cumulative Update[s]” published beginning in 2000 by Ford Bubala & Associates and now carried on by Dr. Sartore, this article (and several of my

Nevertheless, tethered to its roots in competitively proximate marks—(i) *critical to support aiding in the Squirt format and (ii) enhancing the likelihood that both marks will be attended to*¹⁸³—an evolved two room Squirt format utilizing the following elements approaches an Eveready in reliability¹⁸⁴:

- a. multiple and contextual stimuli;
- b. balanced questions with don't know options and admonitions as to guessing; and
- c. controls to filter guessing that nonetheless occurs.

The Eveready and Squirt formats are, indeed, two sides of the same coin. In an unaided Eveready, “respondents conduct a (similarity) comparison between an allegedly infringing mark and **internally accessible information in memory** as to a senior mark.”¹⁸⁵ In an aided Squirt, “respondents conduct a (similarity) comparison between an allegedly infringing mark and a competitively proximate senior mark that is **externally available in the real-world**.”¹⁸⁶

In confusion factor terminology, both formats address similarity of the parties' marks, the *sine qua non* of likelihood of confusion.¹⁸⁷ The Eveready format (**as a probe of memory**) typically relies on the strength factor (the accessibility of trademark information in memory) to facilitate an *internal* similarity assessment. The Squirt format relies on the proximity factor (**the external competitively proximate existence of both marks in the marketplace**) for an external similarity assessment.

The external marketplace, of course, constantly evolves—e.g., salespersons (at least to assist me) in the purchase of goods have substantially vanished¹⁸⁸—and the Squirt format must keep pace in *measuring* whether “an allegedly infringing [competitive] use [is] sufficiently close in time or space to a senior use so that both are sufficiently still in (non-degraded) memory/[cognitive workspace] so as to permit their comparison.”¹⁸⁹ A case in point is *Joules Ltd. v. Macy's Merchandising Group, Inc.*,¹⁹⁰ where plaintiff offered young

prior articles) would not have been possible.

¹⁸³ *Ever-Constant Eveready*, *supra* note 6, at 675-76.

¹⁸⁴ Second Edition, *supra* note 7, at 72-73.

¹⁸⁵ *Id.* at 72.

¹⁸⁶ *Id.*

¹⁸⁷ See Barton Beebe, *An Empirical Study of the Multifactor Tests for Trademark Infringement*, 94 Cal. L. Rev. 1581, 1600, 1607 (2006); see also Daryl Lim, *Trademark Confusion Revealed: An Empirical Analysis*, 71 Am. Univ. L. Rev. 1285, 1329 (2022).

¹⁸⁸ As I have noted in the context of genericism, as a consequence of a changed marketplace, it no longer matters what word a consumer might *use* to ask for what s/he wants when goods are substantially plucked from shelves, and predominant understanding has supplanted use in a genericism assessment.

¹⁸⁹ *Cognitively Updated*, *supra* note 6, at 744.

¹⁹⁰ No. 15-CV-3645 KMW, 2016 WL 4094913 (S.D.N.Y. Aug. 2, 2016).

women's dressy/casual clothing under the JOULES mark through: (i) large national and smaller independent retailers; (ii) its website; and (iii) a few online retailers¹⁹¹; and defendant vended private label MAISON JULES clothing of the same style, price and quality at Macy's and on macys.com.¹⁹²

As to competition, the district court acknowledged that both brands "target[ed] women in their twenties."¹⁹³ As to *physical* proximity, however, it noted that the "products [were] not sold in . . . the same brick-and-mortar stores,"¹⁹⁴ and held that plaintiff's Squirt line-up methodology did not:

accurately reflect the circumstances in which consumers encounter the JOULES and MAISON JULES marks in the marketplace . . . , *i.e.* in the same store or even on the same shelf. See *THOIP [v. Walt Disney Co.]*, 690 F. Supp. 2d [218,] 235 [(S.D.N.Y. 2010)] ("[A] sequential presentation of the two marks at issue . . . is appropriate only if it reflects a significant number of real world situations in which both marks at issue are likely to be evaluated sequentially or side-by-side.").¹⁹⁵

The Second Circuit agreed that the sequential presentation of the stimuli in plaintiff's survey was "artificial[]," particularly as to clothing: a significant percentage of plaintiff's sales through brick-and-mortar outlets were of rainwear, not of work- and weekend-appropriate clothing.¹⁹⁶ The Second Circuit disagreed, however, with the district court's assessment that the proximity factor in a Polaroid analysis weighed "clearly" in defendant's favor: "[b]oth product lines target and are sold to young women who seek work[-] and weekend-appropriate clothing and who are likely to visit both sets of [competing] sales locations [i.e., Joules' retail stores and *website* and Macy's retail stores]"¹⁹⁷ for information or to fill their clothing needs.

In testing whether two marks are likely to be in cognitive workspace for a similarity comparison, a Squirt, to iterate, assesses competitive proximity in the *external* marketplace, as contrasted to

¹⁹¹ *Id.* at *3.

¹⁹² *Id.*

¹⁹³ *Id.* at *8.

¹⁹⁴ *Id.*

¹⁹⁵ *Id.* at *9.

¹⁹⁶ *Joules Ltd. v. Macy's Merch. Grp.*, 695 F. App'x 633, 637-38 (2d Cir. 2017).

¹⁹⁷ *Id.* at 637. Work and weekend wear is a salient category for young women, as were cartoon t-shirts for "involved" grandmothers whom I postulated (contrary to the *Thoip* side-by-side mandate) could carry images of Disney and competing products in their minds from store to store while "on a mission" to find exactly the right shirt for a grandchild. *Cognitively Updated*, *supra* note 6, at 743. My qualms about the *Thoip* mandate applied to salient products that I expressed in 2016 have been effectively resuscitated by internet reality.

an Eveready's *internal* probe of memory. Given constant marketplace evolutions in a free enterprise economy, the full history of the Squirt format may thus never be fully written.

The Second Edition: (i) was largely prompted by the co-editors' recognition that internet surveys have come to dominate data collection and that online shopping has dramatically increased, eroding the brick-and-mortar marketplace still largely in sway (at least for me) at the time of the First Edition in 2012;¹⁹⁸ (ii) it contains an entire chapter devoted to "Internet Surveys in Trademark Cases: Benefits, Challenges, and Solutions"; and it (iii) is interspersed with internet commentary.¹⁹⁹ Just as probability samples were replaced when it was appreciated that businesses largely relied on shopping center intercepts for planning purposes, it is now sometimes difficult to locate field interviewing services because businesses (and survey firms) have switched to the internet as being substantially less expensive for the collection of reliable marketplace information. In the long term, as a measure of external marketplace reality, the Squirt format will more often than in the past prove to be an appropriate methodology to test likelihood of confusion.

In *Joules*, the district court was incorrect in deciding that the Squirt format was inappropriate to test products that—while they did not adjacently appear and compete in the bricks-and-mortar world—were competitively proximate in the broader internet marketplace that exists today for goods and services. The Second Circuit was correct in noting that internet sites and retail stores are frequently in competitive proximity in the currently evolved marketplace.

The side-by-side *Thoip* mandate²⁰⁰ thus is history. A more complicated analysis of how consumers acquire information and shop in a bricks-and-mortar+an-internet-world is upon us.²⁰¹ In a word, the external marketplace has substantially **broadened**. "Determinations as to whether aiding in the [Squirt] format reflects a marketplace reality that two lesser-known marks are likely to be in a consumer's cognitive workspace for a [similarity] comparison have embarked on an entirely new journey."²⁰² Even at 84, I am only clicks away from nutritional information, the status of Alabama

¹⁹⁸ Second Edition, *supra* note 7, at 6.

¹⁹⁹ E.g., my use of Charlie Henn's efforts in *Ever-Constant Eveready* in the Second Edition, *supra* note 7, at 74-77, to identify instances of competitive proximity in the internet marketplace.

²⁰⁰ *THOIP v. Walt Disney Co.*, 690 F. Supp. 2d 218, 235-36 (S.D.N.Y. 2010).

²⁰¹ *Ever-Constant Eveready*, *supra* note 6, at 679.

²⁰² *Id.* In reviewing this article, Dr. Shari S. Diamond aptly raised a question as to whether likelihood of confusion is increased as a result of the broader market or decreased as the concomitant result of increased clutter.

football recruiting, and World News. I am closer to product information on Amazon than I am to actual products in Costco.

Perhaps because the Eveready format was so long deemed limited to top-of-mind marks or because the Squirt format was deemed more amenable for lesser-known marks, there has been a constant tension between survey methodologies for marks that are *internally accessible* in memory as opposed to those that are *externally available* in the market.²⁰³ As one consequence, some experts have stretched the Squirt proximity requirement beyond real-world conditions.

In *Simon Property Group L.P. v. mySimon, Inc.*,²⁰⁴ as a leading example, both parties offered (*noncompeting*) shopping information over the internet, and plaintiff sought permission to use as aiding stimuli: (i) its home page (for major malls across the nation) and (ii) defendant's (more modest) home page as a prelude to the same/different company Squirt questions. The court rejected the side-by-side proposal as (i) grossly distorting conditions under which internet users would likely encounter the two web pages and (ii) designed to create demand effects (given that the two services did not appear adjacently in any marketplace).²⁰⁵

In *Kargo Global, Inc. v. Advance Magazine Publishers, Inc.*,²⁰⁶ the court found that some overlap existed in the demographic make-up of the parties' audiences.²⁰⁷ "Kargo's wireless services and Advance's mens shopping magazine[, however,] . . . were engaged in *different businesses* [and] Kargo . . . offered no data or other evidence to support the proposition that prospective consumers were likely [*in the real world*] to encounter Kargo's trademark [or have it in mind] a short time after seeing *Cargo* magazine."²⁰⁸

Indeed, there was *no real possibility* of confusion in either of the two cases. As in *U.S. Hosiery, supra* Part II, the marks were not competitively proximate and were not familiar to an appreciable percentage of consumers in the sense of lying near to the surface of memory waiting to be triggered by a similar junior use. In *Simon Property Group, Kargo Global* and *U.S. Hosiery*, millions of litigation dollars were wasted on claims that could not withstand a simple marketplace analysis.

²⁰³ Dual or dueling Eveready and Squirt formats are not infrequent. See, e.g., *IDV N. Am., Inc. v. S & M Brands, Inc.*, 26 F. Supp. 2d 815, 821-22, 831 (E.D. Va. 1998) ("Who do you think makes Bailey's cigarettes" produced only 0.2% BAILEYS Irish Cream answers and "[d]o you think there is any connection between [BAILEYS Irish Cream and BAILEY'S Cigarettes]" was deemed leading—plaintiff struck out on both efforts.).

²⁰⁴ 104 F. Supp. 2d 1033 (S.D. Ind. 2000).

²⁰⁵ *Id.* at 1041-49.

²⁰⁶ No. 06 Civ. 550(JFK), 2007 WL 2258688 (S.D.N.Y. Aug. 6, 2007).

²⁰⁷ *Id.* at *7.

²⁰⁸ *Id.* (emphasis added).

As for competitive proximity on the internet, Charles Henn has postulated a (nonexhaustive) need under normal marketplace conditions for:

- (i) a single common search query for the relevant category of products—either on a search engine or on a marketplace site like Amazon or eBay—to produce results in which both sites or products appear on the first page;
- (ii) the use of more than one common search query for the relevant category of products to reveal *each* of the sites or products on the first page of results for at least one such inquiry, so that a shopper attempting multiple common queries is likely to encounter both sites or products within a single shopping session;
- (iii) a consumer to engage in online searching that would expose the consumer to one party's mark and then, shortly in time thereafter, the consumer would physically shop a brick-and-mortar store in a manner exposing the shopper to the other party's mark; or
- (iv) where the parties' products are so directly competitive or overlapping that consumers would be reasonably likely to go directly to the website of the alleged infringer in the course of a shopping process that also involves searching the Internet for other existing sources of the product type (during which the senior user could be encountered).²⁰⁹

As evolved, the Squirt format, like the Eveready format, is comprehensible and is the beneficiary of a plethora of precedent. The Squirt format, however, is not foolproof. In an *internally* focused Eveready study, the respective marks/stimuli are fixed in the sense that: (i) the junior mark must be contextually presented to respondents and (ii) the *internal* content of memory as to the senior mark cannot be manipulated by the “who makes or puts out” question. In an *externally* focused Squirt study, the marks at issue must be contextually displayed, but the additional marks in a lineup or array may be rigged or manipulated to influence a similarity assessment.²¹⁰ A plethora of precedent is available as to manipulations to alert opposing experts and courts to the problem,²¹¹ but the Squirt format, nonetheless, only “approaches” the Eveready format in inherent reliability.

²⁰⁹ *Ever-Constant Eveready*, *supra* note 6, at 682-83.

²¹⁰ See Second Edition, *supra* note 7, at 78 n.111; see also *Black & Decker Corp v. Positex USA Inc.*, No. 11-cv-5426, 2017 U.S. Dist. LEXIS 147463, at *12 (N.D. Ill. Sept. 11, 2017) (staged photo); *Cairns v. Franklin Mint Co.*, 107 F. Supp. 2d 1212, 1219-20 (C.D. Cal. 2000).

²¹¹ See Second Edition, *supra* note 7, at 78 n.111.

In addition, in the aided Squirt two-room format, the senior mark is (i) afforded, at the outset, total awareness, a level that few, if any, Eveready tested marks enjoy; and (ii) one leading survey expert believes that a resulting primacy effect enhances reported confusion levels.²¹² Further, a significant difference may exist between Eveready and Squirt processing of stimuli in making similarity/fit assessments. Extensive research with respect to traditional criminal line-ups establishes that eyewitnesses tend to make relative assessments as to which line-up member most resembles an offender, whereas pattern matching in an Eveready format involves absolute processing between a presented stimulus and a brand schema in memory. Since absolute processing results in fewer false positives and greater overall accuracy,²¹³ it would be preferable for marks in a Squirt line-up to be presented to respondents one at a time to promote absolute processing. Consumers, however, rarely (if ever) see marks one at a time and I have not encountered a decision insisting on sequential displays to enhance the accuracy of Squirt methodology results.

Given my belief that Eveready format results are (somewhat) more reliable than Squirt format results, I postulated in my first article that “net confusion calculations from an Eveready test of less than 10 percent [the threshold noted by McCarthy²¹⁴] should suffice to support a conclusion as to likelihood of confusion (and that, because of an elevated degree of noise . . . , a level above 10 percent should be required from a Squirt format),”²¹⁵ but I could not then and cannot now point to case law or cognitive literature supporting my *ipse dixit*. Whenever, however, there is a battle of experts, one utilizing a properly executed Eveready and the other a properly executed Squirt, I submit that differences in reliability should be one consideration.

A. Subsequent Applications of the Squirt Format

Given its methodological issues, and without a mega brand like MCDONALD’S leading the way, the aided Squirt format crept, rather than leapt, into 20th century. In *Franklin Resources, Inc. v. Franklin Credit Management Corp.*,²¹⁶ for example, respondents were shown (i) four ads, including one by plaintiff, and (ii) three

²¹² See *U.S. Polo Ass’n, Inc. v. PRL USA Holdings, Inc.*, 800 F. Supp. 2d 515, 534-35 (S.D.N.Y. 2011) (“Studies using [Squirt] methodology generally produce higher levels of confusion.” (citing Welter, *supra* note 61 § 6.01[4] (discussing two stimuli Squirt formats as to which higher reported confusion levels is a recognized problem))).

²¹³ Nancy K. Steblay, Jennifer E. Dysart & Gary L. Wells, *Seventy-Two Tests of the Sequential Lineup Superiority Effect: A Meta-Analysis and Policy Discussion*, 17 *Psych. Pub. Pol’y & Law* 99, 102 (2011).

²¹⁴ McCarthy, *supra* note 23, § 32.189.

²¹⁵ *Straitened Scope of Squirt*, *supra* note 6, at 746 n.42.

²¹⁶ No. 95 CIV. 7686(CSH), 1997 WL 543086 (S.D.N.Y. Sept. 4, 1997).

letters, one from defendant, and (iii) were asked if they had seen a product or service from the company whose letter was last read. Sixty percent answered that they remembered the FRANKLIN name, but the court credited criticisms that the survey did not accurately reflect actual market conditions.²¹⁷ The two companies did not compete in the same market and their seriatim presentation was thus artificial.²¹⁸

In *Universal City Studios, Inc. v. Nintendo Co.*,²¹⁹ the question “[t]o 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” was deemed leading.²²⁰ In *Simon & Schuster, Inc. v. Dove Audio, Inc.*, the above noted battle between a Squirt reflecting 36% confusion and an Eveready showing 6.6%, the court found that a closed-ended question was not leading but ambiguous, leaving “no way of knowing” how the respondents interpreted the question.²²¹ The war of experts thus ended in a tie—both surveys were given only marginal value.

In *Sunbeam Corp. v. Equity Industries Corp.*, respondents were shown five food processors, four with a feeding chute and defendant’s compact without a chute and the court remarked that defendant’s processor, “the only one without a distinctive chute, stood out like a bearded man in a lineup with four clean-shaven men.”²²² In *McGraw-Edison Co. v. Walt Disney Productions* (TRON electric fuses v. TRON movie and licensed products), the court noted that the Squirt format “invariably ha[d] a fairly high propor[tion] connecting two unrelated products”²²³ (which, of course, should not be tested using Squirt methodology that placed an electric fuse in artificial proximity to a Disney licensed product).

In *Jordache Enterprises, Inc. v. Hogg Wyld, Ltd.*²²⁴ (JORDACHE designer jeans v. LARDASHE designer jeans for large size women), seventy students were shown the two pairs of jeans and asked their opinions as to their manufacture. Fourteen percent said same manufacturer; and sixteen percent responded that the names were confusing.²²⁵ The district court criticized the side-by-side comparison which did not reflect the reality of the marketplace;²²⁶

²¹⁷ *Id.* at *6-7.

²¹⁸ *Id.* at *1.

²¹⁹ 746 F.2d 112 (2d Cir. 1984).

²²⁰ *Id.* at 118.

²²¹ 970 F. Supp. 279, 290-91 & n.13 (S.D.N.Y. 1997).

²²² 635 F. Supp. 625, 634-35 (E.D. Va. 1986).

²²³ 225 U.S.P.Q. (BL) 512, 517 (N.D. Ill. 1985), *rev'd*, 787 F.2d 1163 (7th Cir. 1986).

²²⁴ 625 F. Supp. 48 (D.N.M. 1985), *aff'd* 828 F.2d 1482 (10th Cir. 1987).

²²⁵ *Id.* at 54.

²²⁶ *Id.*

the Tenth Circuit opined that plaintiff should have used Eveready methodology.²²⁷

From reviewing the case summaries in Welter, *supra* note 61, only toward the latter decades of the 20th century (and confined to the Southern District of New York) did experts begin with some consistency to execute the Squirt format in its current two room/multiple stimuli form. In *Berkshire Fashions, Inc. v. Sara Lee Corp.*,²²⁸ “[s]ubjects . . . were shown a magazine advertisement for a SHEER ENERGY product and then asked to review [an array] of products, including an ENERGIZER slipper” and the Court credited the testimony of plaintiff’s expert that a “28% level of confusion exist[ed] in the marketplace with respect to the SHEER ENERGY and ENERGIZER products in question.”²²⁹ I would have questioned whether the products were competitively proximate (or had been placed in artificial proximity), but the survey otherwise reflected a legitimate two-room execution.

In *Kraft General Foods, Inc. v. Allied Old English, Inc.*:

[R]espondents were first shown a display consisting of either the RAGING BULL Bar-B-Que Sauce or the Texas Best Barbecue Sauce [the control] and 9 other disguise products arranged horizontally on a shelf. They were told to look at each one as if they were considering buying it and were allowed to pick the product up if they desired. All products were then removed from the respondent’s sight. Each respondent was then shown the BULL’S-EYE Barbecue Sauce and four other disguise products, one at a time,²³⁰ and asked whether the product was the same brand, or [was] made by the same company as the (PRODUCT TYPE) she saw earlier.²³¹

A net 26% answered in the affirmative, and in holding that the survey supported an inference as to a likelihood of confusion,²³² the court noted that the methodology was virtually identical to that approved a year earlier in *Kraft General Foods, Inc. v. Friendship Dairies, Inc.*²³³ The Squirt format had finally arrived.

Notwithstanding its slow emergence, a Squirt survey early in the 21st century was appropriately deemed an accepted method in an otherwise erroneous decision in the Ninth Circuit.²³⁴ In *National*

²²⁷ *Jordache Enters., Inc. v. Hogg Wyld, Ltd.*, 828 F.2d 1482, 1487-88 (10th Cir. 1987).

²²⁸ 725 F. Supp. 790 (S.D.N.Y. 1989).

²²⁹ *Id.* at 797-98.

²³⁰ Appropriately, in my view, to promote absolute processing (with fewer false positives) rather than relative processing. See *supra* note 213 and accompanying text.

²³¹ 831 F. Supp. 123, 130-31 (S.D.N.Y. 1993).

²³² *Id.* at 131.

²³³ 19 U.S.P.Q.2d (BL) 1691 (S.D.N.Y. 1991).

²³⁴ *Fortune Dynamic, Inc. v. Victoria’s Secret Stores Brand Mgmt.*, 618 F.3d 1025, 1037-38 (9th Cir. 2010). The district court had rejected plaintiff’s Squirt format survey, but

Financial Partners Corp. v. Paycom Software, Inc.,²³⁵ a properly executed two room Squirt with a control cell reflecting a 39% confusion rate was found to be probative evidence of actual confusion.

B. Variants of the Squirt Format

In my first article, I described variants that I had encountered in litigation as initial efforts at removing the “spotlight from the brands at issue.”²³⁶ Respondents were shown (i) an array or (ii) a sorting board of marks (including the senior and junior uses, a control, and four or five other marks in the category) and were asked:

Do you think [(a)] that each of these brands is from a separate company, or do you think [(b)] that two or more are from the same company or are affiliated or connected [in any way]?²³⁷ If you don’t know, please feel free to say so.

[If (b):] Which two or more brands do you believe are from the same company or are affiliated or connected? . . .²³⁸

In my view, arrays and sorting boards may be appropriate in a grocery store context where a number of brands could be contextually displayed as they exist in the marketplace, but they

because the Squirt format had been approved in *Thane International, Inc. v. Trek Bicycle Corp.*, 305 F.3d 894 (9th Cir. 2002), the Ninth Circuit reversed. In *Fortune Dynamic*, however, plaintiff’s shoes and defendant’s tank tops were not competitively proximate, and the two stimuli test was leading. See 618 F.3d at 1036-37, 1038-39; see also *supra* note 36. The complaint should have been dismissed because the products moved through different channels and the marks were not sufficiently close to the surface of memory so that either would have been triggered upon an encounter with the other. See Zeisel & Kaye, *supra* note 89, at 167-69 (discussing *United States Hosiery Corp. v. Gap, Inc.*, 707 F. Supp. 800 (W.D.N.C. 1989)). See generally *Hi-Tech Pharms. Inc. v. Dynamic Sports Nutrition, LLC*, No. 1:16-cv-949-MLB, 2021 WL 2185699, at *18 (N.D. Ga. May 28, 2021) (where the products “are not sold in the same stores or, for the most part, on the same websites, [the Squirt] format may over-estimate confusion by forcing consumers to consider the marks in close proximity in a way they would not in the marketplace”); *Rex Real Est. I, L.P. v. Rex Real Est. Exch., Inc.*, No. A-19-CV-696-RP, 2020 WL 710198, at *2-3 (W.D. Tex. Feb. 12, 2020) (refusing to strike Squirt format survey since it “is an accepted method” and alleged deficiencies “are properly addressed through cross-examination at trial”); *Superior Consulting Servs., Inc. v. Shaklee Corp.*, No. 6:16-cv-2001-Orl-31GJK, 2019 WL 913374, at *6 (M.D. Fla. Feb. 25, 2019) (Squirt survey showing minimal confusion was “credible and highly persuasive”); *Moroccanoil, Inc. v. Zotos Int’l, Inc.*, 230 F. Supp. 3d 1161, 1175 (C.D. Cal. 2017) (noting that Squirt design is “most appropriate” for goods or services in close proximity).

²³⁵ No. 14 C 7424, 2015 WL 3633987, at *8-9 (N.D. Ill. June 10, 2015).

²³⁶ *Straitened Scope of Squirt*, *supra* note 6, at 749.

²³⁷ In *Scott Fetzer Co. v. House of Vacuums, Inc.*, 381 F.3d 477, 487-88 (5th Cir. 2004), the Fifth Circuit appropriately criticized “in any way” phraseology as prodding “survey participants to search for any connection, no matter how attenuated . . . instead of permitting participants to make their own associations.”

²³⁸ *Straitened Scope of Squirt*, *supra* note 6, at 749.

might be plagued by rigged line-ups, relative assessments and false positives.²³⁹

IV. OTHER METHODOLOGIES

Although the Eveready and Squirt formats were the principal long-term survivors, a number of other methodologies emerged in the latter decades of the 20th century. In *Novo Nordisk of North America, Inc. v. Eli Lilly & Co.*,²⁴⁰ the court rejected a question as to whether plaintiff, the maker of NOVOPEN and other branded insulin injection pens, “had to give its permission or approval” to defendant, the maker of insulin cartridges that were compatible with plaintiff’s pens, for defendant to identify plaintiff’s NOVOPEN and other branded pens on packaging for defendant’s cartridges. It held that “[t]his question mistakenly asks respondents what they believe is the legal requirement (because of the use of the phrase ‘had to’), rather than asking them merely whether they believed that the maker of the [cartridges] did receive authorization to use the names of the pens.”²⁴¹ In *Schieffelin & Co. v. Jack Co. of Boca, Inc.*, a different judge in the same District had held two years earlier that:

Defendants’ objection to the question, “Do you think the company that makes or distributes the product I showed you *had to get* authorization — that is permission — from anyone else to market the product?” as a “legal” question is ineffective. The question sought to reveal whether, upon viewing the CHAMPOP bottle, an ordinary purchaser would believe that the owner of the DOM PÉRIGNON trademarks sponsored or approved their use on defendants’ product — certainly a relevant inquiry under this Circuit’s caselaw.²⁴²

For the reasons stated by Dr. Jacoby, in “Sense and Nonsense in Measuring Sponsorship Confusion,”²⁴³ I agree with the *Schieffelin* ruling. Particularly given the current pervasiveness of licensing,²⁴⁴ I do not believe that a respondent—asked whether a T-shirt company had to get permission to sell a crimson and white shirt emblazoned with a depiction of a red elephant—would feel any need to consult a lawyer before screaming “Roll Tide!”²⁴⁵ The “had to get”

²³⁹ See *supra* note 213 and accompanying text.

²⁴⁰ No. 96 Civ. 5787 (BSJ), 1996 WL 497018 (S.D.N.Y. Aug. 30, 1996), at *6 n.24.

²⁴¹ *Id.*

²⁴² 850 F. Supp. 232, 247 (S.D.N.Y. 1994).

²⁴³ 24 Cardozo Arts & Ent. L.J. 63 (2006).

²⁴⁴ See generally *Bd. of Supervisors for La. State Univ. Agric. & Mech. Coll. v. Smack Apparel Co.*, 550 F.3d 465 (5th Cir. 2008).

²⁴⁵ I apologize for all the SEC references, but as a rabid Alabama fan, they are my best analogies.

terminology “simply taps into knowledge that, as living, breathing members of contemporary society . . . [consumers and] respondents already have.”²⁴⁶ The weight of authority, however, sides with *Novo Nordisk*,²⁴⁷ and given the controversy that it has created I do not believe the “had to get permission” phraseology will again be used.²⁴⁸

In *E.I. DuPont de Nemours & Co. v. Yoshida International, Inc.*,²⁴⁹ the court, citing *Application of Ferrero*,²⁵⁰ held that the “fact that one mark may bring another mark to mind does not in itself establish likelihood of confusion.”²⁵¹ Asked, for example, what “PEPSI brings to mind” a number of respondents might venture COKE (the dominant brand in the cola category that consumers fully appreciate comes from a different company). While “what comes to mind” may be appropriate in a dilution context where association of a mark with a divergent concept²⁵² or “crowding”²⁵³ may be at issue, it triggers (in my view) far too broad a search of memory for likelihood of confusion purposes.²⁵⁴

Nonetheless, in *Exxon Corp. v. Texas Motor Exchange of Houston, Inc.*,²⁵⁵ licensed drivers were shown defendant’s TEXON sign and asked, “What is the first thing that comes to mind?” and “What was there about the sign that made you say that?” Fifteen percent said plaintiff’s EXXON brand and twenty-three percent said gasoline, gas station, or an oil company, which the court (in my opinion erroneously) considered “strong evidence” of likelihood of confusion. A “what makes you think that?” question cannot convert

²⁴⁶ Jacoby, *Trademark Surveys* § 8.41.3 (2013).

²⁴⁷ See, e.g., *Nat’l Football League Props., Inc. v. Prostyle, Inc.*, 18 F. Supp. 2d 1012, 1017-19 (E.D. Wis. 1998); *Louis Vuitton Malletier, S.A. v. Hyundai Motor Am., No. 10 Civ. 1611(PKC)*, 2012 WL 1022247, at *24 (S.D.N.Y. Mar. 22, 2012).

²⁴⁸ “Who sponsors or promotes ____?” can, of course, be used in the Eveready format as an alternative to “who makes or puts out ____?” See *McDonald’s Corp. v. McBagel’s, Inc.*, 649 F. Supp. 1268, 1277 (S.D.N.Y. 1986). And given her credentials (see *Author Biographies of Second Edition, supra* note 7, at xi-xii), I trust that the legal question objection will not be extended to Shari S. Diamond’s question in *Pharmacia Corp. v. Alcon Laboratories, Inc.*, 201 F. Supp. 2d 335, 366 (D.N.J. 2002), “Please tell me whether or not you believe that the company whose advertisement you just saw needs authorization, permission or approval from some other company in order to put out the product advertised.”

²⁴⁹ 393 F. Supp. 502, 519-20 (E.D.N.Y. 1975).

²⁵⁰ 479 F.2d 1395 (C.C.P.A. 1973).

²⁵¹ *Accord Frosty Treats, Inc. v. Sony Comput. Ent. Am., Inc.*, 426 F.3d 1001, 1010 (8th Cir. 2005); see McCarthy, *supra* note 23, § 23:9.

²⁵² See Jerre B. Swann, *Swann’s Response to Diamond Re Dilution surveys*, in *Second Edition, supra* note 7, at 181-82.

²⁵³ Mark A. Lemley, *The Modern Lanham Act and the Death of Common Sense*, 108 Yale L.J. 1687, 1705 n.91 (1999).

²⁵⁴ See *Revlon Consumer Prods. Corp. v. Jennifer Leather Broadway, Inc.*, 858 F. Supp. 1268, 1276 (S.D.N.Y. 1994) (rejecting survey with a question as to whether brands were “in any way associated” since it produced inconsistent, subjective coding of responses).

²⁵⁵ 628 F.2d 500, 507-08 (5th Cir. 1980).

even educated guesses into reliable confusion evidence and I agree with Dr. Itamar Simonson that a “brings to mind” question is a “poor predictor” of likelihood of confusion.²⁵⁶

In *Jefferson Bankshares, Inc. v. Jefferson Savings Bank*,²⁵⁷ 88% of respondents had heard of JEFFERSON SAVINGS bank, but when asked its location, 26% identified a JEFFERSON NATIONAL bank site. In my experience, location studies were done where a large metropolitan bank, say a (fictional) First National of Atlanta, was freed to expand statewide, and opened, e.g., a First National of Hahira, Georgia. The ensuing David versus Goliath battles were intense.

In First Edition, *supra* note 7, at 72-73, I additionally discussed “Going Both Ways,”²⁵⁸ “mystery shoppers,”²⁵⁹ and in-store studies.²⁶⁰ The first has not attracted a following; the second is no longer needed given internet access to niche shoppers; and the third has died from the unwillingness of stores to interfere with customer shopping experiences.²⁶¹

V. REJECTED METHODOLOGIES

In many cases, experts have endeavored to rely on “experience and expertise” (as opposed to data) to support a conclusion that consumers would (or would not) likely be confused as between two marks.²⁶² In every instance that I have encountered, such an effort to supplant the function of the court and a properly instructed panel of “fact finders” has been rejected.

Occasionally, an expert heaves a Hail Mary: “Do you believe [the marks] are likely to produce confusion in the marketplace . . . ?”²⁶³ I wish I knew how here to insert an LOL meme.

VI. CONTROLS

I have appropriately opened and ended my prior discussions of controls with citations to the premier authority, Dr. Shari Seidman

²⁵⁶ Itamar Simonson, *The Effect of Survey Method on Likelihood of Confusion Estimates: Conceptual Analysis and Empirical Test*, 83 TMR 364, 368 (1993).

²⁵⁷ 14 U.S.P.Q.2d (BL) 1443, 1446 (W.D. Va. 1989).

²⁵⁸ See Alex Simonson, *Surveys of Trademark Confusion: Basic Differences*, 5 Intell. Prop. Strategist 1 (1998).

²⁵⁹ See *Lan Tai Shing Co. v. Koch + Lowy*, 19 U.S.P.Q.2d (BL) 1081, at *51-56 (S.D.N.Y. 1990).

²⁶⁰ See Zeisel & Kaye, *supra* note 89, at 156-57.

²⁶¹ Second Edition, *supra* note 7, at 77.

²⁶² See, e.g., *Tovey v. Nike, Inc.*, No. 1:12CV448, 2014 WL 3510636, at *5 (N.D. Ohio July 10, 2014).

²⁶³ *Saxon Glass Techs. Inc. v. Apple Inc.*, 393 F. Supp. 3d 270, 284 (W.D.N.Y. 2019), *aff'd* 824 F. App'x 75 (2d Cir. 2020).

Diamond, “whose long insistence on the utilization of controls”²⁶⁴ in her “Reference Guide on Survey Research,” in the *Reference Manual on Scientific Evidence* (3d ed. 2011), has been the prime factor in their now universal deployment in surveys for trademark litigation.²⁶⁵ I will largely limit my own comments to the observation that unlike the choice of survey formats and stimuli that are largely dictated by cognitive and marketplace considerations, the selection of a control is often a *sui generis* exercise:

As a prime example, Dr. Diamond’s “general principle” that a control “should share as many characteristics with the experimental stimulus as possible, with the key exception of the characteristic whose influence is being assessed” is the best directive we have. At least with respect to multi-element trade dress, however, controls may necessarily incorporate one or more of the characteristics “being assessed” to demonstrate that confusion is being caused, in fact, by the combination [and not by a functional or otherwise non-protectable component of the dress].²⁶⁶

My favorite experience, indeed, with controls was in *Kind, LLC v. Clif Bar & Co.*²⁶⁷ Plaintiff claimed a combination of six elements as its trade dress for a nutritional bar:

(1) packaging with a transparent, rectangular front panel revealing a large portion of the bar itself; (2) a horizontal stripe bisecting the transparent front panel containing the flavor of the bar in text; (3) a text description of the product line . . . in line with the horizontal stripe . . . ; (4) a vertical black band, offset to the side of the package, containing a bulleted list of many of the bar’s key healthful attributes; (5) opaqued . . . end caps . . . ; and (6) a 40g size, in a slender shape.²⁶⁸

Plaintiff conducted a survey testing defendant’s new MOJO bar packaging, which it alleged infringed its trade dress, and used as a control defendant’s prior dress which did not contain any of the six elements; and plaintiff asserted that there was a net confusion of 15%.²⁶⁹ Defendant’s expert testified, however, and the court agreed that:

²⁶⁴ *Cognitively Updated*, *supra* note 6, at 749.

²⁶⁵ Also see her chapters, “Control Foundation: Rationales and Approaches,” in the First Edition, *supra* note 7, at 201-16, and in the Second Edition, *supra* note 7, at 239-55, and Mike Rappeport’s chapter, “Design Issues and the Value of Multiple Controls,” in the Second Edition, *supra* note 7, at 257-79, to which she substantially contributed.

²⁶⁶ *Cognitively Updated*, *supra* note 6, at 750 (footnote omitted).

²⁶⁷ 111 U.S.P.Q.2d (BL) 1795 (S.D.N.Y. 2014).

²⁶⁸ *Id.* at 1797.

²⁶⁹ *Id.* at 1804-05.

[W]hen you have multi element trade dress, a key element . . . is that the multiple elements—all of them . . . collectively could be causing the confusion or individual elements could be causing the confusion. . . . And the problem . . . is that [plaintiff's expert] doesn't address that because his control has none of the elements. So all he is measuring, all he possibly can measure is all or none[.] . . . He has no way of measuring or seeing whether one or two [functional] elements[, e.g., rectangular front panel revealing a large portion of the bar itself,] of the trade dress are the ones causing the confusion²⁷⁰

Similarly, in *24 Hour Fitness USA, Inc. v. 24/7 Tribeca Fitness, LLC*,²⁷¹ the plaintiff's expert rejected defendant's suggestion that "the control should have been another fitness facility that advertised that it was open twenty-four hours a day," because "24 Hour" was a "characteristic at issue."²⁷² The court, however, found the defendant's expert's suggestion "compelling":

[A] number of 24 Hour's enforcement actions have ended in agreements as to acceptable uses of the following words, some as trademarks [or] descriptors: "open 24 hours," "24 hour," "24 hour a day," "Fit 24 Club," "Workout 24/7," "24/7 Tanning and Fitness," and "The 24 Hour Gym." As conducted, the survey does not measure the amount of confusion between "24 Hour Fitness" and names such as these, or a name such as "All Day Gym," for that matter, compared to the amount of confusion claimed with "24/7 Fitness."²⁷³

Dr. Diamond, of course, advocates the use of multiple controls where appropriate,²⁷⁴ and she fully endorses the result in *24 Hour Fitness*. While, therefore, not incorporating the characteristic at issue in a control is a good guideline, it is not an absolute. Perhaps the only absolute is that a control should not be potentially infringing so as to itself generate confusion and the only guidelines that I can otherwise give is a list of cases where controls have been properly deemed to be either good or bad:

- *Facebook, Inc. v. Think Computer Corp.*, 2013 WL 4397052, at *15 n.57 (T.T.A.B. July 23, 2013) (THINKMAIL, combining "think" in applicant's corporate name, with MAIL,

²⁷⁰ *Id.*

²⁷¹ *24 Hour Fitness USA, Inc. v. 24/7 Tribeca Fitness, LLC*, 447 F. Supp. 2d 266 (S.D.N.Y. 2006), *aff'd* 247 F. App'x 232 (2d Cir. 2007).

²⁷² *Id.* at 280.

²⁷³ *Id.* (citation omitted).

²⁷⁴ First Edition, *supra* note 7, at 215-16.

a good control in an Eveready study commissioned by FACEBOOK opposing registration of FACEMAIL);

- *Bruce Lee Enterprises, LLC v. A.V.E.L.A., Inc.*, No. 10 CV 2333(KMW), 2013 WL 822173, at *21 (S.D.N.Y. Mar. 6, 2013) (control T-Shirt, identical to test except filmstrip images were of an Asian male in a martial arts costume, “Bruce L.” wording was removed from hang tag, and Chinese writing was removed from front of the T-shirt);
- *PepsiCo, Inc. v. Pirincci*, 2012 WL 2930650, at *7-8 (T.T.A.B. June 25, 2012) (CAN DO appropriately used as a control for CAN DEW opposed by Mountain Dew);
- *Sara Lee Corp. v. Goldstone Hosiery Co.*, 2001 WL 253616, at *8-12 (T.T.A.B. Feb. 22, 2001) (GOLDEN STEP, GOLDEN SPIRIT, GOLDEN SHEERS, GOLDEN FIT, and GOLDEN TOUCH variously used as controls because L’EGGs is a “familiar” brand);
- *MetLife, Inc. v. Metropolitan National Bank*, 388 F. Supp. 2d 223, 232-33 (S.D.N.Y. 2005) (METROBANK was an appropriate control for METBANK);
- *Skechers U.S.A., Inc. v. Vans, Inc.*, No. CV 07-01703 DSF (PLAx), 2007 WL 4181677, at *9 (C.D. Cal. Nov. 20, 2007) (“Instead of using a control shoe that resembled the Smooth Rider in every respect except its use of a checkerboard design, the study employed a control shoe that differs from the Smooth Rider in a number of significant respects” obviating any causation assessments attributable to defendant’s adoption of a checkerboard design);
- *Makers Mark Distillery, Inc. v. Diageo North America, Inc.*, 703 F. Supp. 2d 671, 693-94 & n.29 (W.D. Ky. 2010) (red wax seal with no drips may itself have been confusing and was not a good control for a red wax seal with drips);
- *Clinique Laboratories LLC v. Absolute Dental, LLC*, 2011 WL 1652171, at *10-11 (T.T.A.B. Apr. 28, 2011) (BOUTIQUE DENTIQUÉ a good control for CLINIQUE DENTIQUÉ).

VII. CONCLUSION

Hopefully, the foregoing will be of assistance to practitioners in choosing and courts in assessing survey methodologies and stimuli selections. Viewing likelihood of confusion studies as (i) *scientific exercises* in (ii) *fairly comparing* (iii) *contextual stimuli* in a manner consistent with (iv) *marketplace conditions* should simplify the process, and judges (whether survey literate or not) are generally attuned to fairness.

Then, with the addition of a few sticks and carrots to promote good practices, any lingering distortions surrounding likelihood of confusion undertakings should languish. I wholeheartedly endorse, e.g., the mention of a “failure to present the products as they were seen in the marketplace, [a flaw that] must have been readily apparent to [plaintiff’s] experienced counsel,” as one element in the award of attorneys’ fees to defendant incurred at trial, after the close of discovery.²⁷⁵ I am likewise heartened by the court’s award of plaintiff’s survey costs of \$20,600 where the court relied on the surveys in adjudicating the case and specifically stated that “the market surveys entered into evidence suggest actual confusion.”²⁷⁶

As for critiques, which are now commonly adduced as to likelihood of confusion surveys, I write in Second Edition, *supra* note 7, at 383-84, that:

Many cases analyze surveys in meticulous detail. Some cases castigate survey experts and counsel who offer what they should know to be defective survey results, even to the point of imposing attorneys’ fees on counsel who proceed in disregard of survey results that do not support the propositions they urge. Few opinions [, however,] grapple with the nuances of critiques, and none take counsel seriously to task when a critique is substantially or totally unjustified. The ledger . . . should be more balanced.

. . . .

If courts do not balance the ledger, they [may over time re]invite [*ipse dixit* “fatal flaw” diatribes of the type that were once commonplace and which (in my opinion) may have led to a decline in the influence of survey evidence in some federal courts²⁷⁷ [see my discussion of *Board of Supervisors for Louisiana State University Agricultural and Mechanical College v. Smack Apparel* in Second Edition, *supra* note 7, at 372-75].

²⁷⁵ *Juicy Couture, Inc. v. L’Oreal USA, Inc.*, No. 04CIV7203(DLC), 2006 WL 2591478, at *6 (S.D.N.Y. Sept. 11, 2006).

²⁷⁶ *Tri-Star Pictures, Inc. v. Unger*, 42 F. Supp. 2d 296, 307 (S.D.N.Y. 1999).

²⁷⁷ *See, e.g., Frehling Enters., Inc. v. Int’l Select Grp., Inc.*, 192 F.3d 1330, 1341 n.5 (11th Cir. 1999) (“This Circuit . . . has moved away from relying on survey evidence.”); *FCOA LLC v. Foremost Title & Escrow Servs. LLC*, 57 F.4th 939, 956 (11th Cir. 2023) (“While [plaintiff] produced two expert witnesses who used internet *surveys* as the bases of their opinions that there was the potential for actual confusion, the District Court discounted this evidence. The District Court stated that *survey* evidence was of slight weight and viewed unfavorably in the Eleventh Circuit. . . .” (emphasis added) (citation omitted)).

**WHY ASK WHY?
A CRITICAL ASSESSMENT OF
AN HISTORICAL SURVEY ARTIFACT**

*By R. Charles Henn Jr.**

TABLE OF CONTENTS

I. Introduction.....	773
II. The Psychology Literature.....	773
III. Likelihood of Confusion Surveys.....	776
IV. What's the Harm in Asking?.....	778
V. Commentators Nudging Us in the Right Direction	780
VI. So, Then, Why Ask Why?.....	782
VII. Not All Open-Ended Questions Are Problematic.....	785
VIII. Conclusion	786

* Trademark Litigation Partner, Kilpatrick Townsend & Stockton LLP, Atlanta, Georgia, Member, International Trademark Association. The author is grateful to Jerre Swann, Dr. Shari Diamond, Hal Poret, Dr. David Neal, Dr. Matthew Kugler, and Dr. Michel Tuan Pham, all of whom provided valuable insights on drafts of this article.

I. INTRODUCTION

Trademark infringement turns on the question of likelihood of confusion; that is, whether the allegedly infringing mark causes consumers to be confused as to the source, sponsorship, approval, or affiliation of a particular product or service. Many trademark litigants support their case with the testimony of expert witnesses who implement surveys of relevant consumers to empirically test that causal hypothesis. Reliable likelihood-of-confusion surveys employ what is known as an “experimental design,” in that they present survey participants with the allegedly infringing mark in a “test” cell and then separately measure the influence of a non-infringing mark in a “control” cell. That test-and-control experimental design allows experts to isolate the impact of the allegedly infringing mark on the surveyed population and, thereby, offer an opinion concerning whether a likelihood of confusion exists.

For at least half a century, psychologists have known that “[p]eople often cannot report accurately on the effects of particular stimuli on higher order, inference-based responses,”¹ and, “[i]t therefore would be quite impossible for them to describe accurately the role played by these stimuli in influencing their responses; and any subsequent verbal report by subjects about the cause of their responses would be at least partially in error.”² Yet, virtually every survey expert will include some form of a “why do you say that?” question in likelihood-of-confusion surveys. Some experts and courts continue to rely (erroneously, in my view) on answers to those questions to determine the critical issue of whether the allegedly infringing mark is likely to cause confusion.

This article explores and analyzes relevant psychological research, the format of questions commonly asked in confusion surveys, and ultimately seeks to answer the question “Why ask why?” Perhaps it is finally time to abandon—or at least to limit undue reliance on—this unnecessary artifact of history and instead rely on the test-and-control experimental design itself to assess whether the allegedly infringing mark is likely to cause confusion.

II. THE PSYCHOLOGY LITERATURE

Likelihood of confusion surveys are designed to answer an empirical question: Does the allegedly infringing trademark cause a likelihood of confusion among the relevant population? Social science researchers regularly assess the causal effects of stimuli in an effort to answer empirical questions, using what is referred to in

¹ Richard E. Nisbett & Timothy D. Wilson, *Telling More than We Can Know: Verbal Reports on Mental Processes*, 84 *Psych. Rev.* 231, 233 (1977) (hereinafter Nisbett & Wilson (1977)).

² *Id.* at 240.

the literature as an “experimental design.” The fundamental structure of experimental design is the use of a *test* condition and a *control* condition to isolate the particular element whose effect one wishes to measure. John Stuart Mill expounded on the basic concept of experimental design centuries ago:

If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance save one in common, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or cause, or an indispensable part of the cause, of the phenomenon.³

Today, experimental design is widely used in pharmaceutical drug trials, in which the *test* condition involves giving one group of patients the real drug and the *control* condition involves giving another group of patients a placebo, which is identical to the real drug except with regard to the active ingredient being analyzed. For example, the color and shape of the pill in each condition is identical and the inactive ingredients in each condition are identical, so the only difference—both perceptually and chemically—is the new active ingredient being assessed. The efficacy of the drug can thus be determined by measuring the difference in participants’ health outcomes between test and control conditions.

The same concept has been incorporated into modern trademark infringement surveys, with the allegedly infringing trademark or allegedly misleading language being isolated through a test-and-control experimental design. “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.”⁴ By following this methodology, experts are able to conclude with reasonable scientific certainty that the difference between test and control conditions is associated with the allegedly infringing mark.⁵ And that difference permits experts to opine as to

³ John Stuart Mill, *A System of Logic*, Vol. 1 (1843) 455 (London: John W. Parker). See also Marilyn B. Brewer & William D. Crano, *Research Design and Issues of Validity in Handbook of Research Methods in Social and Personality Psychology*, 11, 13–14 (H. T. Reis & C. M. Judd, eds., 2014) (“To establish the causal relationship between two variables unequivocally, variation in the causal factor has to be produced or observed under conditions that are isolated from third factors that may produce a spurious correlation. These third variables must either be held constant or uncorrelated with variations in X. This is the essence of the logic of good experimental design.”).

⁴ Shari Seidman Diamond, *Control Foundations: Rationales and Approaches* (hereinafter *Control Foundations*) in *Trademark & Deceptive Advertising Surveys: Law, Science, & Design* 239, 248 (Shari Seidman Diamond & Jerre B. Swann, eds., 2d ed. 2022) (hereinafter *Trademark Surveys* 2d Ed.).

⁵ See *id.* at 247–48 (noting that the certainty of the conclusion is, of course, “[w]ithin the limits of sampling error”).

causal effects, namely that the accused mark *causes* a likelihood of confusion or is *likely* to mislead relevant consumers.

A primary scientific benefit of the test-and-control experimental design is that it avoids a widely acknowledged psychological problem with people's inability to accurately self-report causal reasons for their behavior. In a groundbreaking article published in 1977, Nisbett and Wilson demonstrated through multiple studies that although we can show (through experimental design) that stimuli are having a causal effect, people cannot reliably or accurately "introspect" and explain *why* they reacted to the stimuli in that way. The authors explained that "whatever capacity for introspection exists, it does not produce accurate reports about stimulus effects, nor does it even produce reports that differ from predictions of observers operating only with a verbal description of the stimulus situation."⁶ They further observed:

[W]hen people are asked to report how a particular stimulus influenced a particular response, they do so not by consulting a memory of the mediating process, but by applying or generating causal theories about the effects of that type of stimulus on that type of response. They simply make judgments, in other words, about how plausible it is that the stimulus would have influenced the response.⁷

Based on their research findings, Nisbett & Wilson concluded that "[t]he accuracy of subjective reports is so poor as to suggest that any introspective access that may exist is not sufficient to produce generally correct or reliable reports."⁸ Over the intervening years, psychologists repeatedly have assessed this issue through further research, and the published results have been consistent. People are not able to accurately explain why they react to stimuli—in large part because their higher-order mental processes are not consciously accessible to them.⁹

Instead of accurately identifying the cause of their reaction, people will essentially *manufacture* a plausible post hoc justification

⁶ See Nisbett & Wilson (1977) at 247–48.

⁷ See *id.* at 248.

⁸ See *id.* at 233.

⁹ See, e.g., Petter Johansson, Las Hall, Sverker Sikstrom, Andreas Olsson, *Failure to Detect Mismatches Between Intention and Outcome in a Simple Decision Task*, 310 *Sci.* 116, 118–19 (2005) ("participants may produce confabulatory reports when asked to describe the reasons behind their choices"); see also Timothy D. Wilson & Elizabeth W. Dunn, *Self-Knowledge: Its Limits, Value, and Potential for Improvement*, 55 *Ann. Rev. Psych.* 17.1, 17.7 (2004) ("[M]ental processes are unconscious because of the architecture of the mind, rather than because of repression or suppression [A] large part of mental functioning is [thus] inaccessible to conscious awareness, no matter how much people introspect"); Timothy D. Wilson & Jonathan W. Schooler, *Thinking Too Much: Introspection Can Reduce the Quality of Preferences and Decisions*, 60 *J. Pers. & Soc. Psych.* 181, 182 (1991) ("[P]eople are often unaware of exactly why they feel the way they do about an attitude object.") (hereinafter Wilson & Schooler).

in an effort to provide researchers an answer to the “why did you do that?” question.¹⁰ “When they reflect about their reasons, they thus focus on explanations that are salient and plausible. The problem is that what seems like a plausible cause and what actually determines people’s reactions are not always the same thing.”¹¹ For these reasons, “it may be quite misleading for social scientists to ask their subjects about the influences on their evaluations, choices, or behavior.”¹²

III. LIKELIHOOD OF CONFUSION SURVEYS

The two most widely accepted survey methodologies for assessing likelihood of confusion¹³ are known by the marks at issue in the cases in which they first arose—*Eveready*¹⁴ and *Squirt*.¹⁵ Generally speaking, the *Eveready* survey methodology can be used in most trademark litigation contexts, while the *Squirt* methodology is limited to those in which the parties’ marks are competitively proximate in the real-world marketplace.¹⁶ From their very first use, however, both formats incorporated a question designed to elicit an introspective post hoc verbal report.

The original *Eveready* surveys showed participants the allegedly infringing products and asked one of two sets of questions:

Question 1: Who do you think puts out these mini-bulbs?

Question 2: What makes you think so?

¹⁰ Nisbett & Wilson (1977) at 249 (“When subjects were asked about their cognitive processes, therefore, they did something that may have felt like introspection but which in fact may have been only a simple judgment of the extent to which input was a representative or plausible cause of output.”).

¹¹ See Wilson & Schooler at 182.

¹² *Id.* at 247. To be clear, what is misleading is not *asking* the “why” question; it is interpreting the answer as if it were a genuine indication of the underlying process. Accord Kimberlee Weatherall, *The Consumer as the Empirical Measure of Trade Mark Law*, 80 Mod. Law Rev. 57, 78 (2017) (“Even a consumer acting consciously is not in control of, and cannot access through introspection, any unconscious pre-processing of information: in other words, asking a consumer *why* they have acted in a particular way will not elicit an accurate answer.”) (Emphasis in original.)

¹³ For a detailed discussion of these methodologies, see Jerre B. Swann & R. Charles Henn Jr., *Likelihood of Confusion Surveys: The Ever-Constant Eveready Format; The Ever-Evolving Squirt Format*, 109 TMR 671 (2019) (hereinafter *Ever-Constant Eveready*).

¹⁴ *Union Carbide Corp. v. Ever-Ready Inc.*, 392 F. Supp. 280, 292 (N.D. Ill. 1975), *reversed by Union Carbide Corp. v. Ever-Ready Inc.*, 531 F.2d 366 (7th Cir. 1976).

¹⁵ *SquirtCo. v. Seven-Up Co.*, 628 F.2d 1086, 1089 n.4 (8th Cir. 1980).

¹⁶ See *Ever-Constant Eveready*, 109 TMR at 677. In that article, we used the term “competitive proximity” to refer collectively to the concepts of *physical* and *temporal* proximity that have been used to justify use of a *Squirt* design. Regardless of the terminology, it is well established that for the *Squirt* methodology to be appropriate, the parties’ marks must be encountered in the real-world marketplace in a way that both marks coexist in the consumer’s cognitive workspace simultaneously. *Id.* at 675–76.

Or

Question 1: Who do you think puts out the lamp shown here?

Question 2: What makes you think so?¹⁷

Likewise, the original *Squirt* questionnaire (employed only a few years after *Eveready*) included a “why” question. Survey participants were asked two questions:

Question 1: Do you think SQUIRT and QUIRST are put out by the same company or by different companies?

Question 2: What makes you think that?¹⁸

Thus, in both the original *Eveready* and *Squirt* surveys, Question 2 asked participants to introspect and then report on *why* they believed the allegedly infringing product came from the party they chose as their answer.

The genesis of this “why” question can likely be found in the fact that neither of the original *Eveready* nor *Squirt* surveys included a control condition (i.e., one in which participants answered the same questions but instead saw a control stimulus that did not include the allegedly confusing mark). In the absence of a proper control condition to isolate the marks at issue, perhaps the surveys’ designers believed that the “why” question would support an inference that the answer to the source-confusion question (Question 1) was driven by the similarity of the marks. Regardless of the reason, the “why” question was included then and has become a fixture in virtually all likelihood-of-confusion surveys since.

Meanwhile, in the intervening decades—and in addition to continuing to rely on the “why” question in surveys, survey experts in trademark cases adopted an experimental design in likelihood of confusion surveys, including controls that isolate the mark (or trade dress) being tested.¹⁹ As Dr. Shari Diamond succinctly explained: “By adding one or more appropriate control groups, the survey expert can test directly the influence of the stimulus.”²⁰ Thus, one might assume that once “the tide turned,”²¹ experts would jettison the “why do you say that?” (or “what makes you think that?”) question and instead rely on the control condition, but that assumption would not be entirely correct.²² Even today, most

¹⁷ *Union Carbide*, 392 F. Supp. at 292.

¹⁸ *SquirtCo.*, 628 F.2d at 1089 n.4.

¹⁹ *See Control Foundations*, at 240.

²⁰ Shari Seidman Diamond, *Reference Guide on Survey Research in Reference Manual on Scientific Evidence* 359, 398 (3d ed. 2011).

²¹ *See Control Foundations*, at 240.

²² *See* Jerre B. Swann, *Likelihood-of-Confusion Surveys* in *Trademark Surveys* 2d Ed., 59, 66 n.46 (“[W]ith the advent of experimental designs, ‘why’ questions are no longer necessary, and it would be wholly improper under current standards to give ‘why’ answers causal significance.”).

likelihood-of-confusion surveys—whether based on *Eveready*, *Squirt*, or some “modified” version thereof—continue to employ some version of the “why” question.²³

Survey evidence is the bailiwick of expert testimony. As such, its admissibility in litigation is governed by Federal Rule of Evidence 702, which requires that “the testimony is the product of reliable principles and methods.” Pursuant to the Supreme Court’s decision in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, the trial court must exercise a “gatekeeping” role and be satisfied of the scientific reliability of the expert’s testimony.²⁴ “[I]n the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error.”²⁵

As discussed previously, the psychological literature demonstrates that survey participants’ answers to “why” questions are unreliable as a measure of what *causes* likely confusion²⁶ because they rest on a flawed assumption that people have valid and reliable introspective access to the causes of their judgments. It seems odd, then, that the “why” questions persist—not only in the survey design itself, but also in expert testimony and court opinions that presumably have made it past the reliability “gate” of *Daubert*.

IV. WHAT’S THE HARM IN ASKING?

Despite the unreliability of self-reporting by survey participants to the causal “why” questions, courts continue to reference answers to the “why” questions. Often, courts do so in unscientific ways that compound the ill effects of including this unreliable data in experts’ survey reports and testimony.

This is not a new phenomenon. Indeed, the original *Squirt* court counted as “*same company*” responses from participants who had responded to Question 1 as “*different companies*,” but then “demonstrated confusion” in response to the follow-up “why”

²³ See, e.g., *D. H. Pace Co., Inc. v. Aaron Overhead Door Atlanta LLC*, 526 F. Supp. 3d 1360, 1379 (N.D. Ga. 2021) (in context of *Squirt* survey noting that participants were asked “why” they thought Google listings would take them to the plaintiff’s website); see also *Hypnotic Hats, Ltd. v. Wintermantel Enters., LLC*, 335 F. Supp. 3d 566, 595 (S.D.N.Y. 2018) (in *Squirt* survey participants “were asked to describe why they believe that connection exists”); *Combe Inc. v. Dr. August Wolff GmBH & Co.*, 382 F. Supp. 3d 429, 461–62 (E.D. Va. 2019) (noting that, in *Eveready* survey, “the survey gave the participants the opportunity to explain why they selected their answer”); *T-Mobile US, Inc. v. AIO Wireless LLC*, 991 F. Supp. 2d 888, 924 (S.D. Tex. 2014) (noting 31.6% of participants mentioned color of the store in response to a “What makes you think that?” question in an *Eveready* survey).

²⁴ 509 U.S. 579, 592–97 (1993); see also *Kumho Tire Co., Ltd. v. Carmichael*, 526 U.S. 137, 147–49 (1999).

²⁵ *Daubert*, 509 U.S. at 594.

²⁶ See *supra* notes 6–12.

question.²⁷ Based on this (erroneous) inference, the court then *added* those participants to the Question 1 “same company” participants to drive up the percentage of “confusion” in the survey from 13% to 25%. Despite the fact that “Seven-Up strenuously objected”²⁸ to this approach, the Eighth Circuit affirmed on the basis that the district court’s findings were not clearly erroneous.²⁹

Nor is this phenomenon (of misinterpreting the significance of “why” answers) something relegated to the distant past. In 1999, in *Cumberland Packing Corp. v. Monsanto Co.*, the court used responses to the “why do you say that?” question “[t]o assess the reliability of [the survey expert’s] choice of controls and the conclusions he drew from the data.”³⁰ Based entirely on those “why” responses (and the fact that people mentioned aspects of the packaging not alleged to be infringing), the court essentially discarded the vast majority of participants who were, in fact, confused by the stimuli.³¹ In *Sazerac Co., Inc. v. Fetzer Vineyards, Inc.*,³² the court criticized a *Squirt* survey offered by the plaintiff’s expert for a number of reasons, and some of those critiques may well have been valid.³³ But when commenting on the results of the survey, the court evaluated responses to the “why” question *quantitatively*, noting that “only three percent of participants who found any relationship between the two brands said it was because 1000 Stories ‘looks familiar’ or had ‘similar design or style.’”³⁴

Because many courts (and the Trademark Trial and Appeal Board) have looked to “why” answers as a means of supporting their conclusion that the survey’s principal quantitative results are reliable, it is understandable that many experts are reluctant to abandon the “why” question. For example, in *General Motors Co. v. Urban Gorilla, LLC*, the court relied on the quantitative survey result (erroneously citing only the *test* cell result rather than the result after subtracting the “noise” reflected in the *control* cell) and noted that “when asked why [they thought it was a Hummer], the oft recurring response was ‘because it looks like one.’”³⁵ Another example is *Re/Max Int’l, Inc. v. Trendsetter Realty, LLC*, in which the court cited several responses to the “why” question (and

²⁷ *SquirtCo.*, 628 F.2d at 1089 n.4.

²⁸ *Id.*

²⁹ *Id.* at 1091.

³⁰ 32 F. Supp. 2d 561, 572–73 (E.D.N.Y. 1999).

³¹ *Id.* at 574–75.

³² 265 F. Supp. 3d 1013 (N.D. Cal. 2017).

³³ *Id.* at 1026–27.

³⁴ *Id.* at 1027.

³⁵ No. 2:06-cv-00133BSJ, 2010 WL 5395065, at *18 (D. Utah Dec. 27, 2010).

follow-up probe “what other reasons?”³⁶) as further support that the experimental design accurately reflected a likelihood of confusion.³⁷ In a concurring opinion in *Jack Daniel’s Properties, Inc. v. VIP Products LLC*,³⁸ Justice Sotomayor quoted from answers to the “why” questions to support her argument that surveys may be of more limited value in parody cases because the average consumer may misapprehend the parodist’s “need” to obtain permission.³⁹

At least one court criticized experts—and essentially discounted their surveys entirely—for *not* asking a “why” question. In *Urban Outfitters, Inc. v. BCBG Max Azria Group, Inc.*, the parties submitted dueling confusion surveys (from Robert Klein and Michael Rappeport, respectively) in connection with a motion for preliminary injunction.⁴⁰ After analyzing both surveys, the court held that “[i]n the absence of a ‘why’ question (i.e., why did you select that response) . . . the Court finds that neither expert can testify conclusively as to the reasons for the difference in that [confusion] rate.”⁴¹ Based on decades of psychology research, the court in *Urban Outfitters* was undoubtedly wrong in this regard. Yet, court opinions in this vein improperly deter experts otherwise inclined to move beyond including “why” questions in their experimental survey designs, because they want to avoid any risk of having a court exclude their surveys.

V. COMMENTATORS NUDGING US IN THE RIGHT DIRECTION

The leading trademark-survey scholars repeatedly have identified the psychological problem with relying on “why” questions to assess likely confusion. Dr. Diamond explained eloquently in her recent chapter in the Trademark Surveys treatise:

“[R]easons” can be misleading—that is, they may not be reasons at all. . . . The reason why such pseudo-reasons may occur is that participants 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.

³⁶ Including additional probe questions such as “what other reasons” only exacerbates the initial introspection problem. A response to the initial “why” question will elicit the top-of-mind plausible explanation, which itself is unreliable; forcing further introspection through follow-on probes will just cause further post hoc rationalization. Put another way, if people lack the ability to self-report causal higher-order decision-making, asking them to do so multiple times will not improve their answers.

³⁷ 655 F. Supp. 2d 679, 705–06 (S.D. Tex. 2009).

³⁸ 599 U.S. 140 (2023).

³⁹ *Id.* at 164.

⁴⁰ 511 F. Supp. 2d 482, 498–500 (E.D. Pa. 2007).

⁴¹ *Id.* at 500.

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. The best method for determining the cause of a confusion response is a survey-experimental 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 longer present.⁴²

In his seminal *Psychological Foundations* article, the late Dr. Jacob Jacoby noted that “the process of retrieving information stored in memory to interpret new stimuli is not done with conscious deliberateness, but unconsciously and virtually instantaneously, generally within the first two hundred milliseconds after apprehending the incoming information.”⁴³ He also advised:

[P]sychological theory and research suggest that confusion may occur at levels that are not consciously retrievable. This is because not all of the experiences and information stored in our memory are equally or easily accessible by the individual. Further, not all information residing in one’s memory can be retrieved in response to an interviewer’s questions.⁴⁴

Dr. David Neal more recently has criticized reliance on the “why” question, arguing that “[t]he answers to such questions may often be of somewhat limited value, however, because many processes underlying our decision making are nonconscious, thus leading people to be inaccurate when reporting the source of their decisions. . . . Put simply, we can often tell someone *what* we think, while being unable to tell them *why* we think it.”⁴⁵

Finally, the ever-practical Jerre Swann⁴⁶ commented more than a decade ago that “Dr. Shari Diamond is of the opinion (and I concur) that ‘why’ questions may no longer be necessary.”⁴⁷ Yet, he admitted

⁴² See *Control Foundations*, at 249–50. See also Reference Guide at 397–98 (“It is often difficult for participants to identify accurately the source of their impressions.”).

⁴³ Jacob Jacoby, *The Psychological Foundations of Trademark Law: Secondary Meaning, Genericism, Fame, Confusion and Dilution*, 91 TMR 1013, 1034 (2001) (hereinafter *Psychological Foundations*) (citing B. Azar, *Split-Second Evaluations Shape Our Moods, Actions*, 29 Monitor on Psych. 13 (1998)).

⁴⁴ *Id.* at 1045–46. As discussed *supra* in Part I, higher-order processes are generally not retrievable. See Nisbett & Wilson (1977), at 233.

⁴⁵ David T. Neal, *Psychological Considerations in Designing Trademark and False Advertising Survey Questionnaires*, 273 in *Trademark Surveys* 2d Ed. at 281.

⁴⁶ Mr. Swann is a mentor, friend, and co-author who is entirely responsible for my knowledge in the area of surveys. I refer to him as “ever-practical” because he has long had an uncanny ability to distill theoretical concepts into practical application in trademark litigation.

⁴⁷ Jerre B. Swann, *Likelihood of Confusion Studies and the Straitened Scope of Squirr*, 98 TMR 739, 741 n.13 (2008) (hereinafter Swann, *Straitened Scope*).

that he would “continue to insist on their inclusion in studies [he] commission[ed]” because courts continued to look to them.⁴⁸

VI. SO, THEN, WHY ASK WHY?

The psychological research suggests that “why” answers are unreliable measures of causation, that some leading commentators readily acknowledge this problem, and that courts sometimes erroneously rely on those answers in rendering substantive infringement decisions. So, then, why do experts continue to ask participants *why* they responded to stimuli the way that they did?

My review of the literature, case law, and the testimony of multiple survey experts reveals three primary reasons for pervasive inclusion of the “why” question: (1) as a quality-control measure; (2) as a manipulation check in the absence of a control stimulus or group; or (3) for internal validation of the experimental design, including the choice of control stimuli. None of these, however, justifies continuing this potentially misleading practice.

One common reason identified by experts for the “why” question is that by reviewing the participants’ verbatim responses to the question, they can identify people who are not paying attention or taking the survey seriously and thus exclude them from the final sample. For example, if a participant in an *Eveready* survey identifies the plaintiff as the source of the defendant’s product in response to the “who puts this out?” question, but in response to a “why do you say that?” follow-up responds “asdfgh,” the expert/researcher may decide to exclude that participant from the tabulated sample.

Two flaws exist in this approach. First, better methods exist to ensure the quality of responses and the attention of participants in surveys. For example, high-quality panel providers monitor participant behavior and regularly “purge” from their panels individuals who provide non-quality responses; attention-checking questions (i.e., questions solely designed to weed out participants who are not being careful) are regularly employed in confusion surveys; and experts can review response data and exclude participants who engage in “straightlining” (i.e., claiming to be likely to purchase every product in a disparate list) or monitor time-to-complete data and exclude “speeders.”⁴⁹ The combination of a few or all of these methods—in an experimental design including a

⁴⁸ *Id.*

⁴⁹ For a detailed discussion of various quality-control methods, see Matthew B. Kugler and R. Charles Henn Jr., *Internet Surveys in Trademark Cases: Benefits, Challenges, and Solutions*, 291 *Trademark Surveys* 2d Ed. at 300–306. See also James D. Abbey & Margaret G. Meloy, *Attention by Design: Using Attention Checks to Detect Inattentive Participants and Improve Data Quality*, 53 *J. Operations Mgmt.* 63, 68–70 (2017).

control condition—more than suffices to ensure the reliability of confusion survey data without asking a separate “why” question.

Second, experts often do not actually use “why” questions *exclusively* for quality control and, instead, report verbatim responses as substantive support for their ultimate conclusion (whether it be likely confusion or not). If it were *solely* to be used for quality control (i.e., solely to identify and exclude participants giving gibberish responses), I would have no objection to its inclusion. But once the expert begins relying on the verbatim responses as substantive support for a *causal* conclusion (i.e., quantitatively reporting the trademark-related responses as a basis to prove or disprove likely confusion), the expert is running afoul of established social science research showing that no such conclusions can be drawn.

Another possible justification for the “why” question is that it could serve as a “manipulation check” in the survey. A manipulation check is essentially a question within an experiment designed to assess whether a participant *accurately* perceived the *intended* message of the stimuli presented (i.e., did she process that the stimulus said the cost of the soap was \$1 less than the price of the competitor’s soap?).⁵⁰ As discussed, the introspective “why” question is a fundamentally unreliable measure of *causation*. To the extent that it was used in the original *Eveready* or *Squirt* questionnaires because those surveys lacked any control condition, that justification no longer exists because modern (admissible) confusion surveys use a test-and-control experimental design.⁵¹

The third justification for including the “why” question in a confusion survey is that it purportedly provides some internal validation of the results of the experimental design. In other words, if a number of verbatim responses to the “why” question call out the senior user’s mark (or the junior user’s confusingly similar mark), then that shows (i) that the experimental design “worked” and/or (ii) that the control stimulus appropriately isolated the mark at issue.⁵² These arguments all flow from a flawed premise—namely, that the “why” answers *themselves* are reliable. They are not.⁵³

⁵⁰ Frederick J. Gravetter & Lori-Ann B. Forzano, *Research Methods for the Behavioral Sciences* 177–178 (2018).

⁵¹ See *Control Foundations*, at 240.

⁵² See, e.g., *Straitened Scope*, at 741, n.13 (“information developed from ‘why’ questions may be helpful to counsel in analyzing the case or supporting the efficacy of the control stimulus”).

⁵³ This problem commonly occurs in confusion surveys involving trade dress comprising multiple elements. When asked the “why” questions, participants often will identify only one element of the claimed trade dress or will offer generalized “because of how it looks” explanations. The experimental design *itself* allows the expert to calculate the likelihood of confusion caused by the combination of trade dress elements through the use of a control that modifies those elements. Asking the “why” question (and even attempting to code those as referring to the trade dress) only complicates (or undermines) the

While “why” answers cannot reliably demonstrate causation, they can serve a role in trademark confusion surveys, albeit in a more limited fashion than discussed previously. In an *Eveready* survey, when exposed to an allegedly infringing mark and asked to identify its source, participants will search their memory and identify the mark based on its similarity to what is readily accessible.⁵⁴ Swann has thus argued that “why” answers in *Eveready* surveys provide relevant *content* (rather than quantitative) information concerning whether there is a “fit” between the “stored knowledge” of a well-known senior mark in memory and the similar junior mark.⁵⁵ This is commonly referred to as “pattern matching” or “pattern recognition.”⁵⁶

For participants who reference the senior mark (in an *Eveready* survey), when asked why they believe the junior mark comes from the senior source, one can conclude that the senior mark is readily accessible in memory and that those participants are making a pattern-matching response to the stimulus. In the *Squirt* context, the “why” question is unnecessary for this purpose because the senior user’s mark is presented to the survey participant within the experiment to ensure that the mark is necessarily accessible in memory when the participant encounters the junior user’s mark.

The Board appropriately recognized the validity of this approach in *Facebook, Inc. v. Think Computer Corp.*, when it cited with approval the quantitative results of an *Eveready* survey conducted by the late Dr. Gerald Ford. In that case, “[t]he responses to the followup [sic] questions—‘why do you say that?’—made specific reference to the similarities between the names FACEMAIL and FACEBOOK.”⁵⁷ Dr. Ford and the Board thus used the “why” answers in an appropriately limited fashion (i.e., as evidence of pattern matching by participants) without using them quantitatively to prove causation (i.e., as the basis for Dr. Ford’s opinion that the FACEMAIL mark was likely to cause consumers to be confused).

experiment’s *actual* result with unreliable and incomplete explanations based on post hoc introspective rationalizations. *See, e.g.,* Gucci Am., Inc. v. Guess?, Inc., 831 F. Supp. 2d 723, 741 (S.D.N.Y. 2011) (noting that it is “unrealistic” to expect “survey participants to be able to parse their thought processes with such a high degree of specificity” to name every element of the claimed trade dress “in response to an open-ended ‘why-do-you-say-that’ question”).

⁵⁴ *See Straitened Scope*, at 744–45.

⁵⁵ Jerre B. Swann, *Likelihood-of-Confusion Surveys*, in Trademark Surveys 2d Ed., 59, 63–64 (hereinafter *Likelihood-of-Confusion Surveys*) (citing D. E. Carlston & E. R. Smith, *Principles of Mental Representation*, in Social Psychology: Handbook of Basic Principles 196 (E. Troy Higgins & Arie Kruglansky eds, 1996) (hereinafter Handbook) and E. Tory Higgins, *Knowledge Activation, Accessibility and Salience*, in Handbook at 135).

⁵⁶ *Likelihood-of-Confusion Surveys*, at 63; *see also Psychological Foundations*, at 1037.

⁵⁷ 2013 WL 4397052, at *15 (T.T.A.B. 2013).

Using “why” answers to qualitatively evaluate pattern matching is consistent with the psychology research because the “why” answers are examples of participants searching their memories for the most “plausible” post hoc rationalization.⁵⁸ The “why” question causes introspection—a search of memory to assess whether a mark stored in memory is so similar to the proffered stimulus as to trigger a pattern-matching response. Because a well-known mark is more readily accessible in memory, it is the most plausible explanation for many participants. That said, while “accessibility in memory” may be helpful to an expert seeking to justify her use of the Eveready methodology,⁵⁹ it ultimately is not the question that the fact-finder is considering when evaluating the results of a likelihood-of-confusion survey.

In short, while some value may exist in the “why” question (i.e., as an extra quality-control check, as a means of assessing accessibility in memory, or as a means of evaluating pattern matching), that value is potentially outweighed by the risk of having “bad” data misused or misinterpreted by the fact finder. If survey experts continue to include the “why” question, therefore, it is incumbent on them to explicitly describe how the answers to that question are being used—and how they are *not* being used—in rendering opinions concerning confusion.⁶⁰ Experts who fail to do so risk having courts misapprehend the significance of those answers to the ultimate question of whether and to what extent the allegedly similar mark is likely to cause confusion.

VII. NOT ALL OPEN-ENDED QUESTIONS ARE PROBLEMATIC

The focus of this article has been on the use of a “why do you say that” question specifically in the context of likelihood-of-confusion surveys. In that context, the “why” question asks participants to introspect as to the *cause* of their reaction to the stimuli—something that psychology research shows participants cannot accurately do.

⁵⁸ See Wilson & Schooler at 182 (“When they reflect about their reasons, they thus focus on explanations that are salient and plausible. The problem is that what seems like a plausible cause and what actually determines people’s reactions are not always the same thing.”); *Control Foundations* at 249 (“[R]espondents 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.”).

⁵⁹ See *Ever-Constant Eveready*, 109 TMR at 672.

⁶⁰ The approach of more explicitly and carefully describing the limited relevance and use of the “why” answers is better than the alternative of asking the “why” question but not discussing the answers in the expert’s report. The latter would seem to invite fact-finders to make use of the “why” answer data however they see fit, without guidance from the expert, which could exacerbate the misuse risks identified here.

I am not, however, suggesting that all open-ended questions should be jettisoned from trademark surveys. On the contrary, it is entirely appropriate to ask participants *what their reaction is* to a stimulus (e.g., “what is the main message in this advertisement?” or “what company puts out this product?”). Put another way, a participant can self-report his or her perception; but a participant cannot (or at least *should not*) be asked to self-report on the *cause* of that perception because well-established science shows that such self-reporting as to causation is unreliable.⁶¹

VIII. CONCLUSION

More than a century ago, Dr. Richard Paynter published *A Psychological Study of Trade-Mark Infringement*, in which he argued that psychology had an important role to play in refining trademark law: “Straightforward psychological methods can determine whether the mark complained of does actually cause confusion, what the exact amount of the confusion is, and how it compares with the extent of confusion between other litigated trademarks.”⁶² Looking to the future, Dr. Paynter posited that “the experimental method will aid in perfecting the machinery of law and its administration and in purifying some of the unfair practices in business.”⁶³ More than a hundred years later, we continue to work to “perfect[] the machinery of law” by implementing reliable experiments to measure confusion.

The time has come for lawyers and survey experts to acknowledge what psychology researchers have long known: that people cannot accurately and reliably identify the causal reasons *why* they are reacting to stimuli in an experimental design. Better methods exist to ensure the quality and reliability of survey responses, and lawyers and experts alike should seek to avoid confusing or misleading courts with unreliable data. In short, experts should consider removing the “why do you say that?” or “what makes you say that?” questions in likelihood of confusion surveys, and instead should rely on the test-and-control experimental design to reliably measure the causal

⁶¹ Eddy A. Nahmias, *Verbal Reports on the Contents of Consciousness: Reconsidering Introspectionist Methodology*, 8 *Psyche* 21, 30 (2002) (“[I]t is important, when examining the role of introspection, to distinguish between reports about the contents of conscious states (such as perceptions, beliefs, desires, or emotions) and reports about the causes of those states (such as prior thoughts or experience). To the extent we have direct access to conscious states, it applies to the former, while the latter will involve inferences about which we are often mistaken.”).

⁶² Richard H. Paynter, Jr., *A Psychological Study of Trade-Mark Infringement* 1 (1920). Dr. Paynter was, at the time, a professor of social and abnormal psychology at Long Island University in Brooklyn.

⁶³ *Id.* at 72.

impact of an allegedly infringing trademark on relevant consumers.

GUIDELINES FOR SUBMISSIONS TO *THE TRADEMARK REPORTER*

Founded in 1911, *The Trademark Reporter* (TMR) delivers engaging and comprehensive peer-reviewed scholarship on trademarks and complementary intellectual property from practitioners, academics, and judges worldwide.

The TMR welcomes submissions of articles, commentaries, and book reviews at wknox@inta.org.

The complete TMR Submission Guidelines are available at <https://www.inta.org/resources/the-trademark-reporter/tmr-submission-guidelines/>.

THE TMR ONLINE

International Trademark Association: Issues of the TMR beginning with Volume 105, Number 1, 2015, are available to the public at <https://www.inta.org/TMR>; INTA members may log in for complimentary access to the complete archive.

HeinOnline: Issues of the TMR beginning with Volume 1, Number 1, 1911, are available through HeinOnline.

LexisNexis: Issues of the TMR beginning with Volume 31, Number 1, 1941, are available through LexisNexis.

Westlaw: Issues of the TMR beginning with Volume 80, Number 1, 1990, are available through Westlaw.

PRINTED COPIES

Printed copies of the TMR can be ordered from William S. Hein & Co., Inc. (order@wshein.com).