Eveready and Squirt—Cognitively Updated
Jerre B. Swann

A Plea for the Proper Citation of the Lanham Act
Paul Horton

Commentary: Fashion Dos: Acknowledging Social Media Evidence as Relevant to Proving Secondary Meaning
Ronald Coleman

Commentary: “The Donald” Trumps Claim for Misrepresentations by Licensee
Sheldon Burshtein

Commentary: Descriptive, or Not Descriptive—That Is the Question: A Review Under Turkish Law of Likelihood of Confusion When Trademarks Share Descriptive Terms
Uğur Aktekin, Güldeniz Doğan Alkan, and Zeynep Çağla Özcebe
EVEREADY AND SQUIRT—
COGNITIVELY UPDATED*

By Jerre B. Swann**

I. INTRODUCTION

“Likelihood” of confusion surveys are analogous to “probability” assessments as to whether two objects, A and B, belong to the same class or originate from the same process:

[The probability is] evaluated by the degree to which A is representative of B, that is, by the degree to which A resembles B. For example when A is highly representative of B, the probability that A originates from B is judged to be high. On the other hand, if A is not similar to B, the probability that A originates from B is judged to be low.1

Among the likelihood of confusion factors, similarity of marks is the sine qua non;2 and the above-quoted representativeness heuristic (involving an instantaneous assessment of mark similarities and differences) may inform the probability (or lack thereof) that two marks will be seen as physically or conceptually related.

The majority of likelihood of confusion surveys conducted for trademark litigation are adaptations of either:

---

1. Amos Tversky & Daniel Kahneman, Judgment under Uncertainty: Heuristics and Biases, in Science, vol. 185 (1947) (emphasis added). “Consumers (in addition to or apart from pattern matching[, discussed infra.]) may make computer-quick, automatic assessments of likenesses and differences in drawing inferences [as to whether marks go or belong together].” November 1, 2014, email from Dr. Jacob Jacoby.

2. Barton Beebe, An Empirical Study of the Multifactor Tests for Trademark Infringement, 94 Cal. L. Rev. 1581, 1600, 1607 (2006) (“[t]he similarity of the marks factor is by far the most influential”; “a finding that the similarity of the marks factor does not favor a likelihood of confusion is sufficient to trigger an overall finding of no likelihood of confusion, regardless, it appears, of the outcomes of any other factors).
(i) the *unaided* “Eveready”\(^3\) format, where respondents are shown only the junior mark and search for *similarities* (a “fit”) with any senior brand information accessible in their memory; or

(ii) the *aided* “Squirt”\(^4\) format, where respondents are shown both the senior and junior marks in a manner designed accurately to reflect the reality of the marketplace.

The two formats, neither singly nor in combination, are comprehensive in their coverage. There is, for example, no appropriate format for a senior mark that is: (i) *neither* sufficiently accessible in memory to be cued (in an Eveready) by a monadic exposure to a similar junior use; (ii) *nor* sufficiently proximate to the junior use in the marketplace so that a (Squirt) dual exposure represents the real world experienced by consumers. The gap in survey modalities, however, is not troublesome: if a senior mark is neither stored in memory nor proximate to a junior use, there is little possibility that it can be compared to the junior use sufficiently to generate a likelihood of consumer confusion, and survey evidence cannot ignore that reality.

I have previously discussed (in detail) the Eveready and Squirt formats in an article and book chapter\(^5\) that have been cited in McCarthy and elsewhere.\(^6\) In recent years, I have more deeply immersed myself in the cognitive underpinnings of the two formats, i.e., in how consumers encode, store and retrieve brand information. An understanding of those underpinnings—by courts and practitioners—is essential to an appreciation of the proper uses, and of the limitations (dramatic as to Squirt), of consumer research as to likelihood of confusion issues.

Cognitively, a confusion study entails a *comparison* of marketplace stimuli with existing or aided memory; and a cognitive analysis informs whether a particular comparison has been appropriately implemented, both as to:

(i) the format utilized; and

(ii) the stimuli displayed to respondents.

---

4. Squirtco v. Seven-Up Co., 628 F.2d 1086, 1089 n.4, 1091 (8th Cir. 1980).
5. Jerre B. Swann, *Likelihood of Confusion Studies and the Straitened Scope of Squirt*, 98 TMR 739, 740-54 (2008); Jerre B. Swann, *Likelihood of Confusion, in Trademark and Deceptive Advertising Surveys* 53, 56-71 (Shari Seidman Diamond & Jerre B. Swann, eds., 2012). I have included herein sufficient content from these publications to give this article stand-alone significance, but readers may benefit from a review of the article or chapter for a fuller understanding of these cognitive comments.
The ultimate goal of objectively fair comparisons should guide all likelihood of confusion assessments, and should inform (and simplify) their evaluation.

The goal of this article is to demystify likelihood of confusion survey undertakings by providing a cognitive template as to:

(i) whether an Eveready or Squirt is appropriate, or whether both or neither should be deployed; and

(ii) whether the survey stimuli have sufficient real world currency to support or negate an inference as to likelihood of confusion.

Many likelihood of confusion assessments can be reduced to a two-step process.

II. THE EVEREADY FORMAT

In a typical Eveready study, respondents are shown only a contextual stimulus of the allegedly infringing junior mark, and are asked the open-ended source question, “Who makes or puts [this] out?”, typically followed by “Why do you say that?” In the initial Eveready study, because plaintiff was a substantially anonymous source of Eveready batteries, and defendant’s Ever-Ready lamp was in a different product category, respondents were additionally asked “Please name any other products put out by the same company that puts [this] out.”

Where sponsorship, as opposed to source, is principally at issue, the lead question may be “Who do you believe, if anyone, is sponsoring or promoting ______?” Where the products at issue

7. For reverse confusion, respondents are shown only the senior mark. In proceedings before the Trademark Trial and Appeal Board, where only the right to register (typically in block letters), not the right to use, is at issue, respondents are typically shown the marks in block type on a card that lists the goods or services as to which registration is sought. OMS Investments, Inc. v. Cent. Garden & Pet Co., 2006 TTAB LEXIS 274 (T.T.A.B. 2006).

8. Union Carbide Corp., 531 F.2d at 385-88.

9. Only .6% of respondents identified Union Carbide as the maker of defendant’s Ever-Ready lamps. Id.

10. See McCarthy, supra note 6, § 32:174.

11. 54.6% of respondents answered that the company that put out Ever-Ready lamps also put out batteries. That percentage was particularly telling because the products were not so related that respondents would likely have guessed that an electric lamp maker also made batteries. Had the products been more related, e.g., ham and cheese, then control cell methodology (not in use in trademark studies in 1976) would have been necessary to eliminate “noise.” See, Jerre B. Swann, Judge Richard Posner and Consumer Surveys, 104 TMR 918, 921 (2014). Controls, indeed, are now mandatory irrespective of the relationship between products. See, e.g., Gap, Inc. v. G.A.P. Adventures, Inc., 2011 U.S. Dist. LEXIS 71675, at *24-25 (S.D.N.Y. 2011) (“The Gap” v. “G.A.P. Adventures” for travel services).

12. See James Burrough, Ltd. v. Sign of the Beefeater, Inc., 540 F.2d 266, 278 (7th Cir. 1976). Source, sponsorship, and affiliation are typically addressed in the modern Eveready format.
are in the same category, the “any other products” question may be superfluous. The Eveready format has been so often approved that, beyond these limited variations, deviations may be rejected as lacking objectivity.\textsuperscript{13}

\section*{A. The Cognitive Underpinnings of the Eveready Format}

For a likelihood of confusion to arise, there must be appreciably frequent opportunities for a mental comparison of a senior and an allegedly infringing junior brand. For top-of-mind marks, which are easily accessible in long-term memory, the comparison is accommodated by the Eveready format where respondents are shown only the junior use:

1. Top-of-mind marks exist in memory as elaborate schemas (“clusters” of information) with source-identifying/reputational nodes at their core, strongly linked to the

\textsuperscript{13} See, e.g., McNeil-PPC, Inc. v. Merisant Co., 2004 U.S. Dist. LEXIS 2773, at *59-60 (D.P.R. 2004) where the expert “testified that he was guided by . . . Eveready” in designing the full filter question: “Based on what you just saw, do you or don’t you know who or what brand or company makes or puts out _____?” The “tortured” question, however, substantially reduced “response rate[s],” and the Court, in rejecting the survey, observed that “[t]he actual question asked in \textit{Eveready} was much simpler . . . . The differences are material and likely influenced the responses.”


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 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’. [S]hed in defendants’ shoes, [she] asked the well-to-do passersby whether they could identify the shoes she was wearing. The Court held that “[g]iven the lack of a proper universe and sample, the lack of reproducibility, and the potential bias introduced by the choice of interviewer, . . . this survey in my view has no significant value . . . .” \textit{Id.} at 336.

An interesting Eveready variant is suggested in Hans Zeisel & David Kaye, \textit{Prove It with Figures} 158-59 (1997): to replicate the phenomenon that, as to “frequently purchased household goods . . . [consumers] make quick decisions based on the ‘gestalt’ of the product,” respondents may be shown a notebook with several different pages reflecting experience brands in different categories (one of which pages will include the allegedly infringing junior use), and then be asked to list the brands they recall having seen—with the measure of confusion being those respondents reporting having seen the senior brand.

Another variant—used where the two marks are typically seen in the marketplace side-by-side or where the senior mark is not top-of-mind, but the researcher desires to use the open-ended Eveready questions—is an aided Eveready. “A researcher might present respondents with a store display that includes the products sold side by side (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, which might be called [a] ‘side-by-side Eveready,’ maintains the marketplace reality without asking potentially leading questions.” Itamar Simonson and Ran Kivetz, \textit{Demand Effects in Likelihood of Confusion Surveys, in Trademark and Deceptive Advertising Surveys, supra} note 5, at 249 n.25.
products or services in connection with which they are used and also linked to multiple (usually positive) associations, engrained by advertising, word of mouth and experience.14

2. “These cognitive networks in memory play a fundamental and often decisive role in interpreting incoming information from the outside world.”15

3. Survey respondents, e.g., shown a junior mark and asked “who makes or puts this out?” search their memory and identify the mark “based on its similarity to what [they] already know.”16

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,”17 a process known as “pattern recognition” in which respondents “match” incoming information with a preexisting cognitive network18 (pattern recognition/matching is at the core of an Eveready).

5. Reviews of “why do you say that?” answers often reveal that senior mark responses to the “who makes or puts this out?” question have occurred because “(i) ‘stored knowledge’ of the senior is ‘accessible’ in respondents’ memory, and (ii) there is a ‘fit between that knowledge and the [junior stimulus’”—i.e., that pattern matching has occurred.19

---

14. A highly simplified schema of the ADIDAS mark is as follows:

```
  team focus       Olympics sponsor
  \______________/ \______________/  
 high quality    ADIDAS, three stripes    athletic shoes and wear
  \______________/ \______________/  
    $100+ price point    soccer
```


16. Id. See also Daniel Kahneman, Thinking, Fast and Slow 71 (2011) (“[W]e have in our head a remarkably powerful computer . . . able to represent the structure of our world by various types of associative links in a vast network of various types of ideas.”).


6. “Why?” answers can thus be illuminating, and can confirm, at least in part, respondent associations of a junior brand with a senior.20

B. New Learning from the Consequences of an Eveready’s Underpinnings:
The Format Is Not Appropriate for All Strong Marks

In prior writings, I overemphasized the commercial strength component of a mark as a qualification for Eveready survey treatment, given that strength, among the traditional likelihood of confusion factors, is generally regarded as potentially giving a

20. See, e.g., Audemars Piguet 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 . . . 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%). . . .”); Facebook, Inc. v. Think Computer Corp., 2013 TTAB LEXIS 436, at *42 (“36.2% of the relevant universe. . . expressed a belief that . . . FACEMAIL . . . is either offered [or authorized by or affiliated with] Facebook.”); PepsiCo, Inc. v. Jay Pirincci, 2012 TTAB LEXIS 261, at *18-21 (“all of the 125 confused respondents gave answers indicating that the word Dew was the reason for their confusion [of Can Dew and Mountain Dew].”); Clinique Labs. LLC v. Absolute Dental, LLC, 2011 TTAB LEXIS 161, at *39 (T.T.A.B. 2011) (“. . . each series of the survey questions contain the follow-up question: ‘Why do you say that?’ The answers given to these follow-up questions persuade us that the respondents were not merely reading the stimulus card, guessing or trying to please the interviewer.”); Fiji Water Co., LLC v. Fiji Mineral Water USA, LLC, 2010 U.S. Dist. LEXIS 109155, at *30 (C.D. Cal. 2010) (“24.3% of respondents answered that the VITI water bottle shown to them was made by FIJI or the same company that makes FIJI water. . . . Many of the respondents’ reasons for thinking the product was associated with FIJI were the elements of the FIJI trade dress.”); H-D Michigan, Inc. v. Top Quality Service, Inc., 496 F.3d 755, 758 (7th Cir. 2007) (why answers suggested that confusion stemmed from defendants “use of the word ‘hog’ and that word’s association with the Harley Owners Group”).

Because respondents can sometimes have difficulty accessing their higher-order processes, Richard E. Nisbett & Timothy D. Wilson, Telling More Than We Can Know; Verbal Reports on Mental Processes, 84 Psychol. Rev. 231 (1977), pattern matching is not always evident and “why” answers can never serve as controls to establish causation. Why questions, indeed, may produce “pseudo-reasons [because] respondents who are questioned about their reasons for an answer will search for a plausible explanation that may or may not be the reason for their 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.” Shari Seidman Diamond, Control Foundations: Rationales and Approaches, in Trademark and Deceptive Advertising Surveys, supra note 5, at 211 (citing Nisbett & Wilson, supra note 20). The Court in Mary Kay, Inc. v. Amy L. Weber, 601 F. Supp. 2d 839, 849 (N.D. Tex. 2009) thus erred in using why answers to determine the legal relevance of particular responses. The Court in Sketchers U.S.A., Inc. v. Vans, Inc., 2007 U.S. Dist. LEXIS 88635, at *27-27 (C.D. Cal. 2007), on the other hand, was correct in rejecting an attempt to use “design” or “style” references in why answers, to establish that a checkerboard design caused respondent confusion, given that the “control was useless in eliminating explanations for respondents’ confusion other than the checkerboard design.” See also the discussion of Kind LLC v. Clif Bar & Co., 2014 U.S. Dist. LEXIS 81097 (S.D.N.Y. 2014) in Part VI below.
mark “reach” to a variety of goods.\textsuperscript{21} Cognitively, however, commercial strength, as typically understood, is not sufficient in and of itself; it is the “top-of-mind” characteristic that truly gives a mark reach. “Top-of-mind” refers to marks that are readily accessible in memory. An easily accessible mark effectively \textit{travels in the mind of the consumer}, and reaches—is generally available for \textit{pattern matching} purposes—\textit{whenever} a similar junior mark is encountered (often irrespective of product category).

Strong marks for high-involvement goods that cater to specific and salient consumer preferences and for which there is a not-infrequent need (e.g., athletic shoes) are, almost by definition, top-of-mind. They enable consumers to make instantaneous decisions between products (often with unobservable differences) rather than requiring an examination of ingredients labels or other information to inform a purchase. Commercial strength and top-of-mindness thus frequently cohere in the high-involvement, experience goods category, and the category lends itself to the utilization of the Eveready format.

Conversely, strong marks (measured in terms of sales and market share) for low-involvement goods, such as routinely purchased grocery items (like bread, which is picked from a shelf based on \textit{recognition} of a previously purchased label) may not enjoy high levels of unaided recall because there is no need for consumers to keep them top-of-mind—they are reliably waiting on the shelves of every supermarket. Similarly, a strong mark for rarely purchased, expensive niche items, as to which a consumer will likely canvas the market before making a purchase (e.g., stainless steel grills) might enjoy substantial niche share and sales, but may not be retained top-of-mind (such retention being unnecessary because the mark is easily \textit{recognized} once the search process is initiated).

In each of the latter two instances, the use of an Eveready alone to test for likelihood of confusion may prove problematic. As an example, although I own three grills, I do not carry the luxury VIKING brand in my head because I know I will see and recognize it in any store handling premium grill products. If I thus see “Via-king” on a wok, it is possible that I might not pattern match it with the VIKING grill line; and an unaided Eveready, using the labelled wok alone as a stimulus, might not record me as a confused respondent. If, however, the wok exists in the marketplace in proximity to senior VIKING-branded products, an

\textsuperscript{21} “A mark that is strong . . . is more likely to be remembered and . . . associated in the public mind with a greater breadth of products, than is a mark that is weak because it is relatively unknown.” \textit{James Burrough}, 540 F.2d at 276.
aided Squirt would likely elicit a response from me supporting an inference as to sponsorship confusion.22

The potential ambiguity/difference in my personal assessment as to “who makes” the Via-king wok is largely attributable to the separate cognitive mechanisms that are involved:

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 [and aided] recognition questions generally do a more thorough job of assessing what a person has stored in memory.23

The Eveready format is thus the gold standard for assessing confusion as to (readily recalled) top-of-mind marks;24 but not all commercially strong marks are cognitively stored top-of-mind; and the Eveready format is thus not appropriate for all strong marks. As discussed below, when marks exist proximately in the market, Eveready may not, by itself, be appropriate to test for likelihood of confusion as to strong marks that are generally recognized, but not readily recalled.

That leads to the question, “How top-of-mind—what degree of recall or accessibility—does a mark need to possess to qualify for Eveready analysis?” As of her last edition ofTrademark Surveys,

22. There are undoubtedly many other strong marks that will not generate sufficiently high Eveready percentages to support a conclusion as to likelihood of confusion. In, for example, Kreation Juicery, Inc. v. Eiman Shekarchi, 2014 U.S. Dist. LEXIS 180710, at *19 (C.D. Cal. 2014), a niche mark (found to be strong by the Court) for Mediterranean food and juices made with locally grown, organic ingredients for health conscious diners did not score well among potential purchasers of Mediterranean food generally in an area “with a population of 600,000 people, only a small fraction of whom could reasonably be expected to be aware of [recall] Kreation from prior experience based on the volume of Plaintiff’s business and its targeted appeal to health conscious consumers.” A driving factor of this cognitive update, therefore, is to limit brand “strength,” divorced from top-of-mindness, as an all-purpose qualification for Eveready “Gold Standard” treatment. Only if KREATION had been top-of-mind among “health conscious diners,” who formed a significant subset of Mediterranean diners generally, would an Eveready among the subset have properly and usefully tested likelihood of confusion.


Phyllis J. Welter posited that a brand must enjoy unaided awareness to support the Eveready format. As with most trademark questions, however, the likely more accurate answer is, “It depends”:

Two basic variables influence the likelihood that some stored knowledge will be activated—the accessibility of the stored knowledge prior to stimulus presentation and the fit [or match] between the stored knowledge and the presented stimulus. . . . The greater the overlap between the features of some stored knowledge and the attended features of the 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.

Since both (i) accessibility and (ii) fit impact the percentage of senior mark responses as the “source” of a junior mark stimulus, there can be no fixed level as to either as a prerequisite for use of the Eveready format. Even though, for example, I may not have unaided awareness of the VIKING brand, if I saw VIKING on a wok in the senior brand’s script, a slight variant of the aforementioned example, I am confident that its more precise “fit” would trigger the senior VIKING mark. As I have earlier noted, an Eveready resembles to some degree a partially aided awareness test: Is the junior use sufficiently similar to the senior mark to cue the latter’s schema? In close cases, “Going Both Ways”—separately measuring mark similarity and senior mark awareness—may prove informative.

C. Confirmed Learning as a Consequence of Eveready’s Underpinnings: The Format May Not Be Used Alone to Disprove a Likelihood of Confusion as to Marks that Are Not Top-of-Mind, but Are Proximate in the Marketplace

For marks that are not readily accessible in memory, an Eveready may substantially underestimate the likelihood of real world confusion. Where marks, for example, are in physical or temporal proximity (that is, with an appreciable degree of

27. Higgins, supra note 19, at 135.
frequency, are seen side-by-side or in close sequence in the marketplace), an Eveready *alone*, exposing respondents *only* to a junior use, may not reflect market (dual exposure) reality and cannot suffice alone to *disprove* a likelihood of confusion\(^{30}\)—and recent judicial observations to that effect are heartening.\(^{31}\) Such a distorted deployment of the Eveready format may, indeed, expose both the survey expert and counsel to adverse comment, either express or implied.\(^{32}\)

In *Akiro, LLC v. House of Cheatham, Inc.*, as an example, the marks at issue were the plaintiff’s MISS JESSIE’S and the defendant’s AUNT JACKIE’S, both for natural and curly hair products. In an evident bit of gamesmanship, the defendant’s survey expert endeavored to justify his choice of the Eveready format to disprove a likelihood of confusion by touting the MISS JESSIE’S brand as “well-known” and “strong”—and the Court accepted that characterization in analyzing the brand strength confusion factor.\(^{33}\) The Court went on to note, however, that “a reasonable juror could conclude that [plaintiff’s] mark is weak”—cognitively putting into play whether the negligible results produced by defendant’s survey were attributable to a lack of fit between MISS JESSIE’S and AUNT JACKIE’S, or to MISS JESSIE’S lack of “strength.” Effectively the Court responded to what it apparently perceived as gamesmanship by giving the plaintiff the benefit of opposing inferences: (i) it treated the expert’s use of an Eveready as an acknowledgement that the plaintiff’s mark was strong, but (ii) refused fully to credit the survey’s negligible results because the plaintiff’s mark may have been weak.\(^{34}\)

It is likely, of course, that the expert in *Akiro* parroted the plaintiff’s own allegations of brand strength, in opting for an Eveready,\(^{35}\) and if he had limited his assessment of the survey to establishing MISS JESSIE’S lack of “reach” (as opposed to

---


34. *Id.* at 335, 340.

35. Strong/famous mark claims in pleadings frequently expose a party to the unaided, more rigorous Eveready survey approach, both in federal court (as in *Akiro*) and in the Trademark Office, *see* Clear Choice Holdings LLC v. Implant Direct Int’l, 2013 WL 5402082, at *10 (T.T.A.B. 2013), but I prefer to see the claimant’s credibility directly questioned rather than to subject it to a survey approach that does not reflect the cognitive reality of the case.
claiming that it disproved a likelihood of confusion), I do not believe the Court should have questioned his effort. Particularly with respect to high-involvement experience goods, an Eveready that produces a low level of confusion may provide evidence that a mark is not top-of-mind—that in a confusion factors analysis, a plaintiff is not entitled to count brand “strength” in its favor.

In addition, had the AUNT JACKIE’S and MISS JESSIE’S brands hypothetically not existed proximately in the market (their degree of overlap was disputed), but, for example, were separately vended through different beauty salons, then Eveready would likely have been—for the reasons stated at greater length infra—the only appropriate survey format, irrespective of the weakness or strength of the underlying marks. An Eveready would have been a market reality exercise, and would not have merited a gamesmanship response.

Where, in sum, marks are not proximate, an Eveready is likely the only format available; and even as to proximate marks, an Eveready may always be offered (possibly in response to complaint allegations of trademark strength) in conjunction with a Squirt to negate the reach capabilities of a brand. An Eveready alone, however, cannot persuasively negate an inference of a likelihood of confusion as to marks that are not top-of-mind, but are proximate; just as the below-discussed Squirt format cannot be offered to support an inference of a likelihood of confusion as to marks that are not proximate.

**III. THE SQUIRT FORMAT**

In *Squirtco*, respondents first heard radio ads for SQUIRT and QUIRST soft drinks and were then asked “Do you think SQUIRT and QUIRST are put out by the same company or by different companies?,” followed by “What makes you think that?” Several courts have found such closed-ended questions to be leading. See, e.g., People’s United Bank v. Peoplesbank, 2010 U.S. Dist. LEXIS 60173, at *23 (D. Conn. 2010). When used in connection with only two stimuli, closed-ended questions are strongly

---

36. I keep italicizing “proximity” because it is the *sine qua non* of the Squirt format. See, e.g., Denimafia Inc. v. New Balance Athletic Shoe, Inc., 2014 U.S. Dist. LEXIS 27541, at *96-97 (S.D.N.Y. 2014), where the Court approved an Eveready alone to disprove likelihood of confusion as to (i) not-top-of-mind, but (ii) noncompeting marks, quoting the cognitively sound decision in *Kargo Global, Inc. v. Advance Magazine Publishers, Inc.*, 2007 U.S. Dist. LEXIS 57320, at *23-24 (S.D.N.Y. 2007): “A survey [as to marks that ‘appeal to different audiences in different sectors of the market’] that utilizes the Eveready format, by displaying only a single party’s mark and attempting to discern whether respondents are confused as to the source of the mark, is much more reliable because it more accurately approximates actual market conditions by insuring that respondents are not made artificially aware [as they might be in an aided Squirt] of the other party’s trademark.”

37. See Gosmile, Inc. v. Levine, 769 F. Supp. 2d 630, 643-44 (S.D.N.Y. 2011) (no participant in the Eveready named the not-top-of-mind GO SMILE mark for whitening teeth as the source of GLO tooth whitening products; in the two-stimuli Squirt, 37.5% in the test and 38.5% in the control said the same company put out both products).

38. Several courts have found such closed-ended questions to be leading. See, e.g., People’s United Bank v. Peoplesbank, 2010 U.S. Dist. LEXIS 60173, at *23 (D. Conn. 2010).
currently executed Squirts, questions as to “sponsorship” and “affiliation” typically follow.40

Showing a respondent only two stimuli and asking three two-pronged, closed-ended questions can generate, cognitively, extreme demand effects (skewing “same company” responses substantially higher);41 and while a robust control may temper a traditional Squirt’s suggestiveness, any format that can generate a level of “noise” as high as 80%42 is inherently suspect. Today, therefore, traditional Squirts are infrequently utilized:

the fair and non-leading way in which experts now conduct this type of survey is to show the plaintiff’s and defendant’s product in the context of a number of products about which they would be questioned. This removes the spotlight from the products of the plaintiff and defendant, helps avoid making obvious what the survey is about, and makes the survey more realistic and less leading.43

suggestive and should be avoided, see note 42 infra, but when properly constructed, the better view is that they are not leading. See Jacoby, in Trademark and Deceptive Advertising Surveys, supra note 5.

39. 628 F.2d at 1089 n.4. As an interesting side-note, SQUIRT was deemed a strong mark; in Union Carbide, the District Court (erroneously) found that EVEREADY was descriptive and did not merit protection. The use of Eveready for strong marks and Squirt for weak marks thus represents a role reversal wrought by history.

40. See, e.g., Kargo Global, 2007 U.S. Dist. LEXIS 57320, at *14. As noted in Swann, Likelihood of Confusion, in Trademark and Deceptive Advertising Surveys, supra note 5, at 58 n.34, some courts have expressed concern with the cumulative impact of three closed-ended questions as to “same company,” “affiliated company” and “authorization,” but if the questions are properly balanced and if the expert uses a robust control, the concern, in my view, is often overstated.


42. See Kargo Global, 2007 U.S. Dist. LEXIS 57320, at *26 (“[T]he mere putting of [a closed-ended] question [in a traditional two-stimuli Squirt format can create] the impression of a relationship.”).

43. Tokidoki, LLC v. Fortune Dynamic, Inc., 2009 U.S. Dist. LEXIS 65665, at *21-22 (C.D. Cal. 2009) (plaintiff’s “survey was essentially an artificial matching game in which respondents were shown multiple sets of two products in succession and were asked whether those two products were somehow connected”). The Southern District of New York in (i) Kargo Global (a leading 2007 case with respect to survey analysis that is quoted in notes 36 and 42 supra) and the Tenth Circuit in its leading (ii) Water Pik, Inc. v. Med-Systems, Inc., 726 F.3d 1136, 1148-49 (10th Cir. 2013) decision (“[b]y suggesting the possibility that SinuSense might be connected with another brand, and limiting the candidates to SinuCleanse and NeilMed [a control], the survey questions risked sowing confusion between SinuSense and SinuCleanse when none would have arisen otherwise”), have effectively endorsed Tokidoki; the Ninth Circuit, erroneously in my view, continues to admit two-stimuli Squirts, Fortune Dynamic, Inc. v. Victoria’s Secret Stores Brand Management, Inc., 618 F.3d 1025, 1037-38 (9th Cir. 2010), because that format was approved in Thane Int’l, Inc. v. Trek Bicycle Corp., 305 F.3d 849, 902 (9th Cir. 2002)—a position tantamount to saying that Evereadys and Squirts without control cells are scientific evidence because that is the way they were originally executed.
As the less frequent of two principal “spotlight” removing variants, respondents (as to marks that exist side-by-side in the marketplace) are shown an array or “sorting board” (including the senior and junior uses) and are asked:

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

[IF TWO OR MORE FROM SAME COMPANY OR AFFILIATED/CONNECTED]

Which two or more brands do you believe are from the same company or are affiliated or connected? [and] Why do you say that?

A second (far more frequently used) variant, a two room/line-up study, “is an attempt to replicate the marketplace process of an advertising exposure to a brand or trade dress, followed by being confronted in the market [before the exposure has degraded in memory] with both similar and differing brands or trade dresses”: (i) in one room, a respondent sees the allegedly infringed brand; (ii) then sees in a second room a line-up of brands typically in the same product category, that includes the allegedly infringed brand.

---


45. With respect to using arrays, a critical difference may exist between Trademark Office and court proceedings. In the TTAB, marks are often required to be tested not in context, but in block letter form, OMS Investments, Inc. v. Central Garden & Pet Co., 2006 TTAB LEXIS 274, at *40-41 (T.T.A.B. 2006), and an appropriate array can be easily arranged. Replicating the marketplace with an array may prove more problematic, see Westchester Media Co. v. PRL USA Holdings, Inc., 103 F. Supp. 2d 935 (S.D. Tex. 1999), and because of the difficulty of showing the typicality of side-by-side presentations, arrays are disfavored. See Jacob Jacoby, Trademark Surveys 7.41.1 (2013). As to sorting boards, see Big Dog Motorcycles, L.L.C. v. Big Dog Holdings, Inc., 402 F. Supp. 2d 1312, 1323 (D. Kan. 2005); Sara Lee Corp. v. Kaiser-Roth Corp., 2004 TTAB LEXIS 146, at *41-43 (T.T.A.B. 2004).

46. See Scott Fetzer Co. v. House of Vacuums Inc., 381 F.3d 477, 487-88 (5th Cir. 2004) appropriately criticizing the “in any way” phraseology as prodding “survey participants to search for any connection. . . .”

47. This basic format, as to only three products, was used in Ecce Panis, Inc. v. Maple Leaf Foods USA, Inc., 2007 U.S. Dist. LEXIS 85780, at *16-17 (D. Ariz. 2007) where, as can often occur with three-stimuli Squirt designs, the results of the (in-treatment) control reduced the test results to a level below that supporting a conclusion as to likelihood of confusion.


49. McCarthy, supra note 6, § 32:177. Because of the potential difficulty of showing that the postulated “marketplace process” occurs with appreciable frequency, the rationale may not, in fact, support the utilization of a Squirt format.
infringing brand; and (iii) is asked whether any products in the line-up “come from the same maker or company as the product I showed you [in the first room or do none of these products come from the same maker or company as the product I showed you in the first room?” Rather than on an immediate juxtaposition of the junior and senior brands, it relies on a sequential, aided brand display to make the allegedly infringed brand mentally accessible for comparison.

A. The Cognitive Underpinnings of the Squirt Format

Only a small percentage of almost two million federally registered marks have (or can hope to have) sufficiently strong memory traces so as to be cued by a monadic exposure to a similar junior use. As a consequence, however, of the recency effect, an otherwise not-top-of-mind mark may be sufficiently fresh in memory to permit a mental comparison and pattern matching with a similar junior use that is encountered in temporal or physical proximity with the senior use. Respondents in the most

50. Care must obviously be exercised in selecting line-up participants in order not, effectively, to highlight the allegedly infringing product. See Brighton Collectibles, Inc. v. RK Texas Leather Mfg., 923 F. Supp. 2d 1245, 1258 (S.D. Cal. 2013) (“A line-up in which only one bag shares the most prominent and eye-catching features—two colors and silver hearts—improperly suggested to the participants that Defendant’s bag was the ‘correct’ answer. Consequently, the survey does not prove actual consumer confusion . . ., but instead tested the ability of participants to pick the most obvious match.”), citing Sunbeam Corp. v. Equity Indus. Corp., 625, 634 (E.D. Va. 1986) (“defendant’s product ‘stood out like a bearded man in a lineup with four clean shaven men.’”). See also, Bobrick Washroom Equip., Inc. v. Am. Specialties, Inc., 2012 U.S. Dist LEXIS 111465, at *48-51 (C.D. Cal. 2012) for a dueling Eveready (4.7% confusion) v. Squirt (38.46% confusion) match-up, where the Squirt line-up contained one of defendant’s ads that was nearly identical to plaintiff’s and three substantially different control ads, leading the court to reject the effort and adopt defendant’s Eveready without a discussion as to whether plaintiff’s mark was top-of-mind; Papa Ads, LLC v. Gatehouse Media, Inc., 2011 U.S. Dist. LEXIS 155496, at *19-20 (N.D. Ohio 2011) (an aided Eveready/Squirt variant where respondents were asked “Which of these companies would confuse you into thinking they were either part of iShopStark.com or actually iShopStark.com? Hey Butler.com; ShopNSpark.com; BuyNDine.com”); Bass Pro Trademarks, LLC v. Sportsman’s Warehouse, Inc., 2008 TTAB LEXIS 16 (T.T.A.B. 2008) (a net 35% level of confusion was disregarded “[b]ecause the only similarity between any of the seven marks in the [Squirt] survey was that petitioner’s mark and respondent’s mark both feature the term ‘Sportsman’s Warehouse’.

51. McCarthy, supra note 6, § 32:177; see Jacoby, in Trademark and Deceptive Advertising Surveys, supra note 5, at 274 with respect to the need for “balance.”

52. Hoyer & MacInnis, supra note 15, at 185 (“you are more likely to remember [and be able to retrieve from memory] what you ate for breakfast this morning than what you ate a week ago because (1) this morning’s information has not yet decayed, and (2) there is much less information interfering with its retrieval.”).

53. Forgetting begins virtually immediately, see, Alan Baddeley, Human Memory: Theory and Practice 169 (1990), “with the most rapid decrement occurring earlier rather than later,” November 1, 2014, email from Dr. Jacob Jacoby, which gives rise to the proximity requirement for pattern matching as to not-top-of-mind marks.
frequently utilized (two room/line-up) Squirt are thus presented with multiple marks—first, typically, with the senior mark followed by a line-up of marks in the same category—and are asked whether any one or more in the line-up come from the same company as the mark first shown.

1. Cognitively (to the extent that it reflects market reality), memory of the senior mark is aided to replicate the experience, e.g., of a respondent’s: (a) seeing the senior mark in the marketplace; (b) storing it in memory; and (c) then encountering the junior mark for comparative purposes, before memory of the senior mark has degraded.

2. To assess the probability—or draw inferences as to the likelihood—of a relationship between marks, respondents may pattern match or resort to the representativeness heuristic, quoted in the opening paragraph supra, to “reduce the complex tasks of assessing probabilities . . . to simpler judgmental operations”; i.e., respondents, as in an Eveready, may search for similarities and differences (a fit or match, or the lack thereof) between aided memory of the senior mark and a mark in the line-up.

54. The category requirement is dictated by (i) the need to avoid demand effects, Itamar Simonson and Ran Kivetz, Demand Effects in Likelihood of Confusion Studies: The Importance of Marketplace Conditions, in Diamond and Swann, supra note 6, at 243; and (ii) the reality that only marks in the same category are likely to be seen in temporal or physical proximity with sufficient frequency to support the aided Squirt format. “[I]n cases where the products do not share the same market, the likelihood of confusion is reduced by the very fact that no consumer would ever be exposed to both products.” Hans Zeisel and David Kay, Prove It with Figures 167 (1997). Quoting from case law, Dr. Jacoby notes that “[p]roximity in the marketplace addresses the extent to which the parties’ products compete with each other.” Jacob Jacoby, Trademark Surveys 3.12.1 (ABA 2013).

55. For likelihood of confusion purposes, of course, it is a matter of indifference that a consumer might actually first see and store in memory the junior mark and then encounter the senior mark before the memory of the junior decays.

56. See supra note 53.

57. A significant difference may exist between an Eveready and a Squirt search for similarities. Extensive research with respect to traditional criminal line-ups establishes that witnesses tend to make assessments as to which line-up member relative to the others most resembles the offender, whereas pattern matching in an Eveready involves absolute, one-on-one processing. Since absolute processing results in fewer false positives and greater overall accuracy, Nancy K. Steblay, Jennifer E. Dysart & Gary Wells, Seventy-two Tests of the Sequential Line-up Superiority Effect: A Meta-Analysis and Policy Discussion, 17 Psychol. Pub. Pol’y & Law 99 (2011), it would be preferable for marks in a Squirt line-up to be presented one-at-a-time to promote absolute processing.

Consumers, however, rarely see sequential presentations in the marketplace, and courts may not sanction them in a survey context, much less insist on them as a precondition for Squirt validity. Controls in a Squirt thus assume an added dimension of criticality—the elimination of noise that results when respondents select a line-up mark “most like” the test mark, rather than one perceived (in an absolute sense) to come from the same company.
The representativeness heuristic (and comparable relatedness heuristic\textsuperscript{58}) are appropriate in assessing the likelihood of whether brands that, in the real world, are frequently encountered in physical or temporal proximity will be seen as so physically or conceptually similar or related that they are deemed to go or belong together. For likelihood of confusion purposes, the representativeness heuristic may thus be viewed as a variant of, or work in conjunction with, pattern matching to identify brand stimuli and to evaluate their relatedness.

\textbf{B. New Learning from the Consequences of a Squirt’s Underpinnings: The Format Involves Aided Pattern Matching in the Mind}

I have previously characterized the Eveready format, cognitively, as pattern matching “in the mind,”\textsuperscript{59} as opposed to a Squirt as “pattern matching in the external marketplace.” As a consequence, however, of Squirt aiding, both formats effectively entail pattern matching (or an instantaneous search for similarities and differences) in the mind with the potential for at least partial self-validation from answers to “why” questions,\textsuperscript{60} with at least two major differences:

1. In Squirt, the senior mark is afforded (temporarily at least) total recognition, a level that few, if any, Eveready tested marks enjoy. This may explain the generally higher confusion levels from Squirt designs,\textsuperscript{61} and suggest that an elevated floor should be applied to Squirt undertakings beyond the suggested 10% floor in McCarthy.\textsuperscript{62}

2. Marks that are not top-of-mind marks do not travel with the consumer and are not likely to reach to non-proximate

\textsuperscript{58} Gita V. Johar & Michel Pham, Relatedness, Prominence, and Constructive Sponsor Identification, 36 J. Mktg. Research 299 (1999).

\textsuperscript{59} Swann, Likelihood of Confusion, in Diamond and Swann, supra note 5, at 77.

\textsuperscript{60} In the days before control cell methodology was common in likelihood of confusion surveys, why answers giving a trademark related reason for a “same company” response were (erroneously) used to validate Squirt results. See, e.g., Sally Beauty Co., Inc. v. Beautyco, Inc., 304 F.3d 964, 979-80 (10th Cir. 2002). In a Squirt, however, such answers are far more likely to be post hoc rationalizations based on what a respondent has just seen, see supra note 20, rather than evidence that a respondent has accessed unaided memory. In my view, post hoc rationalizations were at work among the 17.7% in Kargo Global, 2007 U.S. Dist. LEXIS 57320, supra note 42, at *16-18, who gave the “name(s)” as the reason for their “same company” (80%) responses.


\textsuperscript{62} McCarthy, supra note 6, at § 32:189.
goods; the inclusion of non-proximate goods in the line-up after aiding has occurred may thus afford reach that does not exist in the real world marketplace.

C. Confirmed Learning from the Consequences of a Squirt’s Underpinnings: Proximity Is Paramount

The principal confirmation from a review of Squirt’s cognitive underpinnings is that the tested marks must, with appreciable frequency, exist in physical or temporal proximity in the marketplace so that the aiding in the format is reflective of market reality and a Squirt does not create an “artificial” market.63 The degree of physical or temporal proximity that must exist to support a Squirt will vary, of course, depending on the facts of the case.

I have long been concerned, for example, with the stress on same store, side-by-side proximity as a Squirt prerequisite as is reflected by much of the analysis in THOIP v. Walt Disney Co.64 The THOIP images were sufficiently unique and the cartoon characters on the Disney shirts were sufficiently famous that I believe they would not have rapidly degraded and could have been carried in the memories of consumers from store to store, particularly in the memories of “involved” shoppers (e.g., grandmothers) for t-shirts for small (grand)children.65 Consequently, a balanced presentation of both physical and temporal proximity under circumstances of consumer involvement may have been preferable as an approximation of the market.

I suspect that THOIP, by itself adducing “same store” evidence, may have inadvertently given credibility to Disney’s insistence on side-by-side proximity.66 THOIP might have more profitably utilized a memory expert, who could have testified as to the extent that involvement can extend the retention of images in memory, to legitimize its evidence as to “different stores within various distances of each other,” particularly given the Court’s finding that “the Little Miss THOIP, Little Miss Disney, and Miss Disney lines are within the same narrow category of goods directed at the same set of consumers.”67

It must also be appreciated that in today’s multifaceted and complex marketplace, there are products that, while infrequently seen side-by-side, are (i) complementary and/or (ii) purchased by substantially overlapping consumer groups—that is, that are seen by consumers in sufficiently tight timeframes, sufficiently often, so

64. 690 F. Supp. 2d 218, 237 (S.D.N.Y. 2010).
65. See Swann, Likelihood of Confusion, in Diamond and Swann, Surveys, supra note 5, at 68, n.93.
66. 690 F. Supp. 2d at 233.
67. Id.
as to present realistic comparative opportunities. Of perhaps even greater import it must be appreciated that today’s marketplace includes the Internet and that proximity may exist outside the bricks and mortar context. For products that are frequently bought or researched on the Internet, the proximate appearance of two brands in commonly conducted (v. contrived) GOOGLE searches may well satisfy the proximity proof requirement, and stimuli in such instances should be selected to reflect the Internet environment.68

Indeed, given the explosion of Internet use, a host of creative approaches to proximity may exist: for example, (i) I often decide on a wine purchase as a dinner guest, based on my liking of the host’s offering, and I turn to GOOGLE for source information; and (ii) if figures exist as to the frequency of such purchasing behavior, then cyber-world proximity might serve as a substitute for physical proximity in cases such as Leelanau Wine Cellars, Ltd. v. Black & Red, Inc.69 (where the tested wine, in the real world, was not sold in stores and could not thus be tested with other wines in a Squirt two room lineup).

Concepts of involvement and marketplace complexity cut, of course, in multiple directions: a large supermarket, for example, may include 30,000 SKUs (many of them mundane) spread over 50,000 square feet. As, however, “the straitened scope of Squirt” phrase in my prior article implies, I am strongly of the view that a Squirt proponent must demonstrate that there are “a significant number of real world situations in which both marks at issue are likely to be evaluated in close proximity or side-by-side.”70 The mere fact that two marks are found in cluttered grocery venues does not remotely establish that their aided comparison is reflective of market reality, and just as “different store” evidence does not necessarily negate the propriety of the Squirt format, “same store” evidence does not alone satisfy the proximity imperative. 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—and that consideration is not a fixed concept, but can fluctuate depending on a host of mental and marketing influences.

IV. THE COVERAGE GAP

As noted in the Introduction above, the Eveready and Squirt formats, neither singly nor in combination, cover all likelihood of


70. 69 F. Supp. 2d at 235.
confusion survey scenarios. In *U.S. Hosiery Corp. v. The Gap*, for example, Dr. Hans Zeisel (the father of consumer research for litigation purposes) was faced with a dispute involving U.S. Hosiery’s Workforce mark for socks sold in Sears (the senior mark) and Workforce for blue jeans sold in The Gap.71 U.S. Hosiery had commissioned a side-by-side survey of the marks, and 35% of the sample responded that the two products came from the same company.

Based on the results of two surveys designed to measure actual (as opposed to aided) awareness of the marks and the likelihood that consumers would ever encounter both products, Dr. Zeisel concluded on behalf of Gap that the likelihood of consumer confusion was not 35%, but was less than 1% among both Gap customers and the general population.72 U.S. Hosiery’s mark was: (i) neither top-of-mind nor was it proximate to The Gap’s Workforce brand; (ii) it thus fell into the coverage void; and (iii) there was no appropriate survey format to measure reliably any confusion caused by The Gap. The 35% “same company” response level reported from U.S. Hosiery’s two-stimuli Squirt, like the 80% produced from the artificial market in *Cargo Global*, was virtually all noise and would likely have been fully dissipated by an appropriate control.

As I have previously written, a review of the Eveready and Squirt formats leaves a question “as to how likelihood of confusion can be tested when a weak mark does not appear in proximity to or otherwise overlap a similar junior use”:

>[T]he answer is that confusion cannot be assessed and, under such circumstances, likely cannot exist. *Comparisons in appreciable numbers cannot occur.* A mark that is not top-of-mind cannot be confused with another’s use in a commercial arena where the weak mark does not appear and there is no customer or other commonality.73

To iterate, “for a likelihood of confusion to arise, there must be appreciable opportunities for a mental comparison of the senior and junior brands,” and where neither an Eveready nor a Squirt facilitates or replicates such a comparison *in the real world marketplace*, then neither should be deployed; indeed, the *bona fides* of a confusion claim under such circumstances should be open to challenge without any need for empirical survey support.

---


V. FATAL COGNITIVE FLAWS IN LIKELIHOOD OF CONFUSION STUDIES

McCarthy appropriately notes that the “the closer the survey methods mirror the situation in which the ordinary person would encounter the trademark [or trademarks], the greater the evidentiary weight of the survey results.” The reliability of a survey is greatly reduced or even nullified by an expert’s (1) inappropriate format choice, or (2) creation of a survey market that has no real-world counterpart:

(1) Use, therefore, of an Eveready alone to test, not the reach, but the likelihood of confusion as to two proximate marks under circumstances where an aided Squirt would assess viable (real world) comparison opportunities is a fatal flaw; and

(2) Use of a Squirt for marks that cannot be shown to appear, with frequency, in physical or temporal proximity creates an artificial market for comparison purposes.

Beyond format considerations, it likewise may be a fatal flaw, and preclude a fair comparison, to: (3) present survey stimuli abstractly, without adequate context; or (4) utilize distorted stimuli.

(3) Abstract stimuli presentations. In Water Pik, Inc. v. Med-Systems, Inc., a leading 2013 case with respect to survey analysis, the 10th Circuit noted (in disregarding a senior user’s survey): (a) that “[m]arks should be compared ‘as a whole as they are encountered by consumers in the marketplace’”; (b) that “Questions 3 and 10 [in the survey] did not present the marks as they would appear to a consumer because both marks were in typewritten format, divorced from packaging, and without any italics; and the SinuSense mark was unaccompanied by the distinctive ‘waterpik’ logo that appears on packaging for SinuSense products”; and (c) that “Questions 3 and 10 [thus] exaggerated the similarities between the two marks, likely increasing the confusion of the respondents.”

74. McCarthy, supra note 6, at § 32:163.

75. 726 F.3d 1136, 1145-47 (10th Cir. 2013), aff’g 2012 U.S. Dist. LEXIS 81592 (D. Colo. 2012); accord, Hornady Mfg. Co., Inc. v. Doubletap, Inc., 746 F.3d 995 (10th Cir. 2014).

76. The decision in Water Pik is one in a long line of cases condemning abstract presentations, as frustrating the fair cognitive (contextual) comparison that should underlie all likelihood of confusion assessments. In Water Pik, the presentation “exaggerated” the similarity of the two marks. I have been involved in other cases where the abstract presentation masked the difference in the underlying products. In either event, the frustration of a fair comparison of marks as they appear in the real world marketplace is, in my view, cognitively fatal. See, e.g., J.T. Colby & Co., Inc. v. Apple, Inc., 2013 U.S. Dist.
(4) Distorted stimuli presentations. In Kargo Global, my favorite case from 2007 (given the Court's perceptive analysis of multiple cognitive concepts), the expert effectively tripled-down—he presented sequentially (a) two stimuli (b) that would not normally be sequentially encountered and (c) constructed stimuli to suggest an overlap between (i) on-line and (ii) magazine content that did not realistically exist:

The initial phase of the survey consisted of showing the respondents materials from Cargo magazine[, a men’s shopping magazine]. Each respondent was first informed that he would be shown a cover of a magazine and . . . was then shown the cover from the September 2005 issue of Cargo, which displayed, among other things, the “CARGO” logo, a picture of the tennis player Andy Roddick, and several large-font “headlines” that announced the issue’s content. . . . Included on the cover was a prominently displayed picture of a cellphone, to the side of which was printed the headline “World’s most distinctive Cell Phone.”

* * *
Each respondent was then informed that he would next be viewing a “magazine ad that appeared in Premiere magazine.” The respondent was then shown an advertisement for “Premiere Mobile,” a service which allows cellphone customers to download Premiere’s content onto their cellphones. The advertisement included the words “Powered by KARGO”, with Kargo's logo prominently displayed in large font.

Each respondent was then asked a series of [three] multiple choice questions [as to source, affiliation or sponsorship, e.g.]:

Q4. If you have any thoughts about it, do you think these ads from Premiere Magazine. . . .?

Do come from the same company that put out the magazine you looked at earlier in the survey

Do not come from the same company . . .

Don’t know.77

Eighty percent (in both the test cell and the control cell) responded affirmatively to one of the three questions. A fair cognitive comparison was precluded not only by the use of the Squirt format as to non-proximate marks in different categories, but equally, in my view, by effectively distorted stimuli that substantially exaggerated mark similarities and the slight overlap between the parties’ target audiences.78


78. In addition to cases as to abstract stimuli, which represent one form of distorted stimuli, Kargo Global is only one of a long list of cases condemning other forms of distorted stimuli. See, e.g., Tokidoki, LLC v. Fortune Dynamic, Inc., 2009 U.S. Dist. LEXIS 65665 (C.D. Cal. 2009) (“The leading nature and lack of realism of [plaintiff’s] survey was exacerbated by the use of Tokidoki products that were very atypical (decorated [as were a number of defendant’s products] only with a heart and crossbones image and none of the colorful Japanese-inspired artwork and characters typically found on Tokidoki products) and did not give the respondents the kind of exposure to Tokidoki that they would have if exposed to Tokidoki products in the real world.”); Vista Food Exch., Inc. v. Vistar Corp., 2005 U.S. Dist. LEXIS 42541 (E.D.N.Y. 2005) (rather than comparing the two logos at issue, plaintiff's expert effectively compared “Vista” alone with “Vistar” alone, “without the stylized capital V, the proper font, the proper color (green letters with a yellow star above a stylized V), and the VSA”); 1-800 Contacts, Inc. v. WhenU.com, 309 F. Supp. 2d 467, 499-500 (S.D.N.Y. 2003) (“the survey failed to use any stimulus that would inform consumers as to the competing products or marks in question”), rev’d on other grounds, 414 F.3d 400 (2d Cir. 2005); Waddington North American Business Trust v. EMI Plastics, Inc., 2002 U.S. Dist. LEXIS 16634, at *24 (E.D.N.Y. 2002) (“The trays were shown separate from their packaging at a significant distance, and the conditions appear to have been chosen to make it difficult for the participants to discern EMI’s logo before being asked to identify the tray brand.”); The Learning Network, Inc. v. Discovery Comm’ns, Inc., 153 F. Supp. 2d 785 (D. Md. 2001) (“once the subject reached the portfolio pages purportedly depicting the website, he was denied significant location information that would be available to a real world
Quite simply, both Eveready and Squirt (with controls) are scientific exercises in “pattern matching” and can only produce reliable results if the stimuli deployed fairly reflect the reality of the marketplace. If stimuli are shown to respondents out-of-context or are distorted, they cannot be fairly compared. In Water Pik, the similarity of the marks was exaggerated; in Kargo Global, the degree of market overlap was skewed. In neither case were respondents afforded an objective opportunity to match the stimuli or assess their relatedness.

VI. CONTROLS

Any discussion of likelihood of confusion surveys would be incomplete without an acknowledgement of the crucial importance of controls. Given, however, that controls more often involve “analytic[s]” and “engineering” as opposed to cognitive principles, the issue is best addressed by reference to Dr. Diamond, whose long insistence on the utilization of controls in her “Reference Guide[s] on Survey Research” in the Reference Manual on Scientific Evidence has been a prime factor in their now almost universal deployment in surveys for trademark litigation. I will limit my own comments to the observation that unlike the choice of survey formats and stimuli that are largely dictated by cognitive considerations (e.g., a Squirt as to non-proximate marks creates an artificial market for trademark comparisons), the selection of a control is often a sui generis exercise.

consumer. Subjects were provided with a picture of the website which omitted the ‘title bar’ at the very top of each website page. The ‘title bar’ for the Learning Network home page includes the statement ‘Welcome to the Learning Network—your educational resource.’); Beverage Mktg. USA, Inc. v. South Beach Beverage Corp., 2000 U.S. Dist. LEXIS 16504 (S.D.N.Y. 2000) (“Comparing the bottles, with the labels properly on them, unlike plaintiff’s survey comparison of the bottles with that significant element of their trade dress removed, the Court holds that a reasonable jury . . . could not find a likelihood of confusion. . . .”); Thane Int’l Inc. v. Trek Bicycle Corp., 53 U.S.P.Q. 1523, 1531-32 (C.D. Cal. 1999) (“The pictures employed by [plaintiff’s expert] to conduct [this] survey do not strike this Court as accurately replicating the environments in which consumers would encounter the parties marks. . ..”).


80. Mike Rappeport, Design Issues for Controls, in Diamond and Swann, Surveys, supra note 5, at 224.

81. The need for controls in consumer research, particularly to establish causation, was consistently recognized well before such research became prevalent in trademark cases. Some early proponents of Eveready argued, however, that open-ended questions were unlikely to generate guessing that, as a survey artifact, needed to be filtered from likelihood of confusion results. Early Squirt proponents argued that causation could be demonstrated by counting only trademark relevant responses to “why questions” as evidence of confusion. Both were wrong: brand dominance in an Eveready can lead to Squirt-like, two-stimuli mismeasurements and Squirt “why” answers are often post hoc rationalizations. Largely as a result of Dr. Diamond’s efforts, both camps have come to appreciate their error.
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.

In *Kind LLC v. Clif Bar & Co.*, to illustrate, plaintiff claimed a trade dress that consisted of a combination of:

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. opaque . . . end caps . . .; and
6. a 40g size, in a slender shape.

As a control, plaintiff used packaging with none of the elements (its expert stating that the control would otherwise contain an active ingredient) and its survey produced net confusion of 15%.

For the proposition that confusion was caused by the overall design, plaintiff’s expert initially cherry-picked “why do you say that?” responses that highlighted “packaging similarity.” Defendant, however: (a) pointed to numerous “why” references to “flavor similarities”; (b) noted the ambiguity of packaging references (which could have been triggered by the predominant, but not separately protectable, transparent panel); and (c) tested a separate control with a large transparent panel that produced the same level of confusion as had the plaintiff’s test cell. Defendant’s expert thus contended (and the court agreed) that:

When 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

---


83. 111 U.S.P.Q.2d 1795, 1797 (S.D.N.Y. 2014) (for purposes of assessing potential bias, it should be noted that I was involved in the case, presenting the testimony of Clif’s survey expert witness).
seeing whether one or two of the [functional] elements of the trade dress are the ones causing the confusion...84

Similarly, in *24 Hour Fitness USA, Inc. v. 24/7 Tribeca Fitness, LLC*85 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.” The court, however, found the suggestion “compelling”:

a number of 24 Hour’s enforcement actions have ended in agreements as to acceptable uses of the following words, 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.”86

Dr. Diamond, of course, advocates the use, where appropriate as in *Kind v. Clif*, of multiple controls,87 and she fully endorses the result in *24 Hour Fitness*. She articulates four features as characterizing “a good [well-engineered] control stimulus.”88 Reading the many cases that have found controls either productive89 or deficient90 may shed light on the process, and

84. *Id.*, at 1805.
86. *Id.*, at 280.
88. “1. A good control stimulus shares features with the allegedly infringing mark . . . other than those alleged to be infringing . . . that might affect responses; 2. it does not contain cues that will artificially depress confusion responses by leading the respondent in a different direction; 3. if it is a control for an allegedly infringing mark, it should appear to be a plausible member of the same product category; and 4. if it is a control for an allegedly infringing mark, it should not be an infringing mark.”
89. Facebook, Inc. v. Think Computer Corp., 2013 TTAB LEXIS 436 (T.T.A.B. 2013) (THINKMAIL, combining “Think” in applicant’s corporate name, with MAIL, a good control for FACEMAIL in an Eveready commissioned by FACEBOOK); 1-800 Contacts, Inc. v. Lens.com, Inc., 722 F.3d 1229, 1247 (10th Cir. 2013) (asking respondents to imagine they had searched for “contact lenses,” a good control for asking them to imagine that they had searched for “1-800-contacts,” in assessing source of defendant’s ad); Bruce Lee Enters., LLC v. A.V.E.L.A., 2013 U.S. Dist. LEXIS 31155, at *65-68 (S.D.N.Y. 2013) (control T-shirt identical to test except film strip images were of an Asian male in a martial arts costume, “Bruce L.” was removed from hand tag, and Chinese writing removed from front); First Nat’l Bank in Sioux Falls v. First Nat’l Bank S.D., 679 F.3d 763, 770 (8th Cir. 2012) (quoted Dr. Diamond’s general principle in agreeing with plaintiff that “First Bank & Trust” was not improper as a control for failure to include “First National”); Pepsico, Inc. v. Jay Pirincci, 2012 TTAB NEXIS 261 (T.T.A.B. 2012) (Can Do used as control for Can Dew opposed by Mountain Dew); The Gap, Inc. v. G.A.P. Adventures, Inc., 2011 U.S. Dist. LEXIS 71675 (S.D.N.Y. 2011) (where the G.A.P. logo was intended as an acronym for the phrase “great adventure people,” Great Adventure People Adventures was used as control for
applying Dr. Ivan Ross’s chapter, “The Use of Pilot Tests and Pretests in Consumer Surveys,” may inform the ultimate selection of a control. There are, however, simply no rules that will lead unerringly to the correct result.

VII. CONCLUSION

A likelihood of confusion test thus cognitively entails a comparison or pattern matching of marks, as would typically occur in the real world marketplace, to assess their likenesses and differences—to ascertain the existence of a fit (or the lack thereof)—and it may be:

a. **unaided**, where the senior mark is sufficiently top-of-mind so as to be available in memory for comparison to a monadic exposure to a junior use; or

b. **aided** where the reality of the marketplace is such that one mark is likely to be seen in such temporal or physical proximity to the other so as to be still available in memory for a fit assessment to occur.

The stimuli used in the comparison must likewise comport with marketplace reality—they:

a. must be fairly presented in context and undistorted; and

b. cannot create an artificial market or be arranged so as to suggest a particular response.

---


90. Steak Umm Co., LLC v. Steak ‘Em Up, Inc., 2012 WL 1231875, at *17 (E.D. Pa. 2012) (“Steak’n Bake” or “Steak’n Shake” would have been better controls than “Steak Out” because they look and sound more like Steak Umm); LouisVuitton Malletier, S.A. v. Hyundai Motor Am., 2012 U.S. Dist. LEXIS 42795, at *70 (S.D.N.Y. 2012) (“by depicting the ball [shown to control group] in an atypical color that mimics the Louis Vuitton marks, a reasonable trier of fact may question the control’s credibility”); U.S. Polo Ass’n, Inc. v. PRL USA Holdings, Inc., 800 F. Supp. 2d 515, 532-36 (S.D.N.Y. 2011) (while accepting PRL’s survey as “suggestive of actual confusion, the Court noted that it’s control “could have perhaps shared more features with the test product in terms of the shape and material of the fragrance box”; the Court rejected U.S. Polo’s survey for including “the very elements being assessed”); Maker’s Mark Distillery, Inc. v. Diageo N. Am., Inc., 703 F. Supp. 2d 671, 693-94 (W.D. Ky. 2010) (red wax seal with no drips may have itself been confused and was not a good control for a red wax seal with drips); Tokidoki, LLC v. Fortune Dynamic, Inc., 2009 U.S. Dist. LEXIS 65665, at *21-22 (C.D. Cal. 2009) (“A control should be as close to the allegedly infringing design as possible without itself being infringing. . . . The paisley and cherry print designs . . . as ‘controls’ were so different from the heart and crossbones design that their use could only have picked up some of the noise, not all of it.”); Citizens Banking Corp. v. Citizens Fin. Grp., Inc., 2008 U.S. Dist. LEXIS 36800, at *26-28 (E.D. Mich. 2008) (as a control, CHASE bore very little resemblance to the marks at issue).

91. In Diamond and Swann, Surveys, supra note 5, at 11.
If, therefore, the expert and practitioner: (a) pick a format that facilitates appreciable opportunities for real world comparisons of marks; and (b) then pick stimuli that promote fair comparisons of those marks, they are well down the road to a reliable survey product.

Understanding the cognitive underpinnings of Eveready and Squirt should lead to the correct selection of the format best to facilitate the comparison (or to the conclusion that no appropriate format exists); and understanding the need for a real world comparison should mandate the rejection of abstract and distorted stimuli. Viewing, indeed, likelihood of confusion studies as scientific exercises in fairly and objectively comparing stimuli in a manner consistent with marketplace conditions should simplify the process and may obviate other cognitive concerns.92

Of perhaps greatest importance, judges (whether survey literate or not) are well (innately) attuned to fairness and fair comparisons. Their application of cognitive guidelines in that regard should lead to more predictable outcomes in confusion survey assessments; and that, in turn, may deter efforts to distort the process.

92. E.g., priming.