Likelihood of Confusion Surveys

Jerre B. Swann, William H. Brewster, J. David Mayberry, and R. Charles Henn, Jr.
Likelihood of Confusion Surveys
Jerre B. Swann, William H. Brewster, J. David Mayberry, and R. Charles Henn, Jr.

I. Introduction

In the 2006 Kilpatrick Stockton Desk Reference Book, we covered surveys generally. The majority of surveys under the Lanham Act address likelihood of confusion, and the predominant formats, from the onset, have been Eveready and Squirt. These formats differ dramatically in: (a) their opening questions; (b) the means by which they replicate market reality, access brands, and facilitate inferences as to source; (c) the confusion factors they measure; and (d) the circumstances under which they may provide evidence supporting (or negating) a conclusion as to a likelihood of confusion. This article focuses exclusively on the Eveready and Squirt formats.

In 1976, in Union Carbide Corp. v. Ever-Ready, Inc., the Seventh Circuit endorsed Eveready in litigation involving the EVEREADY mark for batteries. Over time, it has become the gold standard in cases where the senior mark is strong, i.e., highly accessible in memory, enhancing the likelihood that it will be cognitively cued by a junior user’s mark.

In 1980, in Squirtco v. Seven-Up Co., the Eighth Circuit held that results from a Squirt study supported a district court finding as to the “possibility of confusion” between SQUIRT and QUIRST for non-cola soft drinks. Over time, the Squirt format has come to be used in cases where accessibility of the senior mark in memory is low to non-existent, so that it must be made externally available to respondents as part of the survey design.

The analysis below confirms, as to strong marks, the gold standard status of the Eveready format. As to the Squirt format, this analysis explores two frailties: (a) the design uses closed-ended, “suggestive” questions; and (b) it may create an “artificial marketplace” for respondent assessments.
of brand similarity. Thus, a Squirt survey must include a robust control, and must be limited to the conditions of its origin (directly competing or substantially overlapping goods) where the stimuli proximately tested in the format do appear, in fact, proximately in the marketplace.

A. Eveready

1. The Questionnaire and Variants

In an Eveready survey, a respondent is first shown an exemplar, photograph, or advertisement of defendant’s branded (or “trade dressed”) product; then, the respondent is asked an open-ended “source confusion” question: “Who makes or puts out ______?” That question is followed by: “Why do you say that?” Questions as to “sponsorship confusion” and “affiliation confusion,” often in closed-ended form, typically follow, e.g.:

Do you believe that whoever makes or puts out ______: ONE, is sponsored or approved by another company; TWO, is not sponsored or approved by any other company; or THREE,
you don’t know or have no opinion? 16 [If ONE] What other company? [and] Why do you say that? 17

The basic format has been approved so often that material variants are infrequent. 18

The follow-on questions, of course, may be phrased in a more open-ended fashion 19 and, under particular circumstances, a follow-on question may become the lead inquiry. 20 Radical alterations, however, often suggest a departure from the requirement that a design be objective: e.g., the question “based on what you just saw, do you or don’t you know who or what brand or company makes or puts out _______,” followed by an “admonishment” to respondents to answer only if they “knew the correct answer,” was rejected as “tortured” and calculated to “produce a low response rate.” 21

Nonetheless, variant approaches, one of them stemming from Union Carbide itself, do exist, both in the case law and the literature:

(1) Where the owner of the senior mark is substantially anonymous and the defendant’s goods are in a different category, the respondent is asked: “Please name any other products put out by the same concern which puts out . . . .” 22

16 “[P]resentation of an explicit ‘don’t know’ or ‘no opinion’ alternative commonly leads to a 20%-25% increase in the proportion selecting that response.” Diamond, supra note 8, at 250, and a number of courts insist on “don’t know” alternatives. See, e.g., Cumberland Packing Corp., 32 F. Supp. 2d at 572; Procter & Gamble Pharms., Inc. v. Hoffmann-La Roche Inc., No. 06 Civ. 0034 (PAC), 2006 WL 2588002, at *22-25 (S.D.N.Y. Sept. 6, 2006). Choices in closed-ended questions must “cover all possible answers a respondent might give to the question [including “don’t know.” If the list . . . is incomplete, a respondent may be forced to choose one that does not express his or her opinion.” Diamond, supra note 8, at 253, citing Am. Home Prods. Corp. v. Johnson & Johnson, 654 F. Supp. 568, 581 (S.D.N.Y. 1987).

17 In Eveready designs, follow-ons are generally considered as the “standard type and format of questions used to gauge confusion. . . .” Pharmacia Corp. v. Alcon Labs., Inc., 201 F. Supp. 2d 333, 365 (D.N.J. 2002). Failure by a defendant, seeking to disprove a likelihood of confusion, to test for “sponsorship or affiliation” may lead to the rejection of its study. Bear U.S.A., Inc. v. Kim, 71 F. Supp. 2d 237, 252 n.106 (S.D.N.Y. 1999), aff’d, 216 F.3d 1071 (2d Cir. 2000). With Squirt designs, some courts have expressed concern with the cumulative impact of a series of closed-ended questions as to “same company,” “affiliated company,” and/or “sponsorship.” See U.S. West, Inc. v. Hatten Commc’ns Holding Co., No. 110, 126 2002 T.T.A.B. LEXIS 620, at *15–20 (T.T.A.B. Sept. 25, 2002) (unpublished) (“we have accorded no weight to opposer’s survey. Not only are the questions memory based [the stimulus was removed from respondents’ view before the questionnaire was administered], but in addition, a respondent merely had to give a ‘wrong’ answer to one of the pertinent [three] questions in order to be counted as part of the 22.9% of the respondents who were ‘confused.’ ”). The “wrong answer” concern is alleviated in Eveready designs where respondents are usually required to identify the “other company,” and I am unaware of any instance where a makes/sponsors/affiliated trilogy in an Eveready test has engendered the strong demand effects that can surface in a Squirt design. See, e.g., Kargo Global, 2007 U.S. Dist. LEXIS 57320, at *17, where 80% of respondents in both the test and the control gave answers reflecting confusion to a closed-ended trilogy.

18 The most bizarre is reported in Arche, Inc. v. Azaleia, U.S.A., 882 F. Supp. 334, 335 (S.D.N.Y. 1995):

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]hod in defendants’ shoes, [she] asked the well-to-do passersby whether they could identify the shoes she was wearing.

19 For example, in James Burrough Ltd. v. Sign of Beefeater, 540 F.2d 266, 278 (7th Cir. 1976), a precursor to Union Carbide, the sponsorship question was: “Who do you believe is sponsoring or promoting this restaurant?”

20 Id.

21 McNeil-PPC, Inc. v. Merisant Co., No. 04-1090 (JAG), 2004 WL 3316380, at *20 (D.P.R. July 29, 2004) (defendant’s expert “testified that in formulating this question, he was guided by the survey. . . . in Eveready . . . . The actual question asked in Eveready was much simpler . . . . The differences . . . are material, and likely affected the responses.”).

22 Union Carbide Corp., 531 F.2d at 385 n.11. Only 0.6% of respondents identified Union Carbide as the maker of defendant’s Ever-Ready lamp; 54.6% answered, however, that the same concern that put out Ever-Ready lamps also put out batteries, supporting a finding of likelihood of confusion and secondary meaning of EVEREADY as well. Id. at 381.
necessity, they “are highly selective.”

bombarded by brand stimuli and cannot “attend” to all that fall within the range of their senses. a “babel of brands,”

understand current conditions of clutter. Half a century ago, Ralph S. Brown, Jr. wrote that there was

KILPATRICK STOCKTON

KILPATRICK STOCKTON


Ralph S. Brown,

impression on the public . . . .”). With respect to many items, however, consumers can be expected to study packages, and a

details for any appreciable length of time”). With respect to many items, however, consumers can be expected to study packages, and a

stimulus should be left with a respondent.


can have no way of knowing whether or not permission was obtained. Jacob Jacoby, Marketing management

32 F. Supp. 2d 561, 578 (E.D.N.Y. 1999). Critics of “needed to get” wording insist it calls for a legal

158 (it should be noted that Dr. Zeisel’s approach, for point of sale confusion purposes, may be


2. Categorization and Pattern Matching in an Eveready Format

To appreciate fully the benefits of an Eveready approach and variants, it is first necessary to understand current conditions of clutter. Half a century ago, Ralph S. Brown, Jr. wrote that there was a “babel of brands,”26 and the “number of choices has [since] grown dramatically.”27 Consumers are bombarded by brand stimuli and cannot “attend” to all that fall within the range of their senses.28 Of necessity, they “are highly selective.”29

23 See, e.g., Cairns v. Franklin Mint Co., 107 F. Supp. 2d 1212, 1219 (C.D. Cal. 2000), aff’d, 292 F.3d 1139 (9th Cir. 2002). In a Squirt survey, the alternative is “aided.” e.g., did A need to get/get permission from B? Critics of “needed to get” wording insist it calls for a legal conclusion. Nat’l Football League Props., Inc. v. Prostyle, Inc., 57 F. Supp. 2d 665 (E.D. Wis. 1999). Critics of “did get” wording insist that a respondent can have no way of knowing whether or not permission was obtained. Jacob Jacoby, Sense and Nonsense in Measuring Sponsorship Confusion, 24 CARDozo ARTS & ENT. L.J. 63 (2006). As noted elsewhere, we regard the debate largely as a waste of judicial resources. Jerre B. Swann, U.S. Trademark Surveys, in Survey Evidence and the Law Worldwide 333–34 (Lexis Nexis 2008). See Pebble Beach Co. v. Tour 18 I Ltd., 155 F.3d 526, 544 (5th Cir. 1998).

24 HANS ZEISEL & DAVID H. KAYE, PROVE IT WITH FIGURES: EMPIRICAL METHODS IN LAW AND LITIGATION 158 (1977). For how long a respondent should be exposed to a stimulus is a function of consumer involvement in the purchase. Hurried consumers, for example, do not typically study items, and stimuli reflecting such items should not be left with the respondent during the administration of a questionnaire. See Am. Home Prods. Corp. v. Procter & Gamble Co., 871 F. Supp. 739, 748 (D.N.J. 1994) (“consumers do not normally meditate . . . over . . . details for any appreciable length of time”). With respect to many items, however, consumers can be expected to study packages, and a stimulus should be left with a respondent. See Cumberlaid Packing Corp. v. Monsanto Co., 32 F. Supp. 2d 561, 578 (E.D.N.Y. 1999) (“In an actual market situation, the product would not disappear from the consumer’s eye just as he or she is about to make a purchase.”). See generally Jerre B. Swann, A ‘Reading’ Test or a ‘Memory’ Test: Which Survey Methodology Is Correct?, 95 TRADEMARK REP. 876 (2005), and Mike Rappeport, Response to Survey Methodology Articles, 96 TRADEMARK REP. 769 (2006) (advocating the more conservative view that the stimulus, in most circumstances, be left with the respondent). See also Henry D. Ostberg, A Response to an article entitled: A “Reading” Test or a “Memory” Test: Which Survey Methodology Is Correct? 95 TRADEMARK REP. 1446 (2005) (advocating the opposite).

25 ZEISEL & KAYE, supra note 24, at 158–59 (it should be noted that Dr. Zeisel’s approach, for point of sale confusion purposes, may be inappropriate where the junior and senior brands are typically shelved adjacent to one another so that respondents should have the context of the senior brand to make judgments as to the junior, Winner Int’l LLC v. Omori Enters., Inc., 60 F. Supp. 2d 62, 71 (E.D.N.Y. 1999)).

26 Ralph S. Brown, Advertising and the Public Interest: Legal Protection of Trade Symbols, 57 YALE L.J. 1165, 1197 (1948).


28 Jacob Jacoby, The Psychological Foundations of Trademark Law: Secondary Meaning, Genericism, Fame, Confusion and Dilution, 91 TRADEMARK REP. 1013, 1022 (2001) (“the vast majority of stimuli fail to register upon the consumer’s consciousness”); Wayne D. Hoyer & Deborah J. MacInnis, Consumer Behavior 115 (3d ed. 2004) (“Shoppers in a supermarket are exposed to numerous products, brands, ads, displays, signs, prices, logos, and packages all at the same time. We are generally unable to examine all those marketing stimuli simultaneously.”).

29 PHILIP KOTLER & KEVIN LANE KELLER, MARKETING MANAGEMENT 186 (12th ed. 2006) (“It has been estimated that the average person may be exposed to over 1,500 ads or brand communications a day. Because a person cannot possibly attend to all these, most stimuli will be screened out . . . .”). See Toro Co. v. ToroHead, Inc., 61 U.S.P.Q.2d 1161, 1180 (T.T.A.B. 2001) (“Every day consumers are bombarded with hundreds, if not thousands, of advertisements for hundreds of products . . . . A great many of these ads do not make a significant impression on the public . . . .”).
Even when a stimulus registers on their consciousness, consumers “rarely . . . consider all [its] features;” rather, given their attention constraints amid clutter, they use shortcuts to “label, identify, and classify” the information. If, for example, they see a small creature with the salient features of feathers and wings, they “categorize” it as a bird without stopping to test for all avian characteristics.

In memory, strong brands function, in part, as antidotes to clutter; they exist as schemas—as “clusters” of information: (a) with source identifying (reputational) nodes at their center; (b) very strongly linked to the product(s) or service(s) in connection with which they are used; and (c) also linked to (usually a host of) other associations that have been engrafted on the schema by advertising, word of mouth or experience. Cognitively, “[a] unique brand name and cohesive brand identity are probably the most powerful pieces of information for consumers . . ., enabling [them] to efficiently organize, store, and retrieve information from memory.” Strong brands operate, moreover, much in the manner of the picture of a celebrity on the cover of a magazine in a sidewalk kiosk – they attract attention in an otherwise hurried environment.

The adidas schema, as an example, has the brands (e.g., adidas, three stripes) and other indicia (e.g., shell toe, Superstar design) at its core strongly linked to athletic shoes and wear; they are also linked to a perception of high quality, soccer, sponsorship of the Olympics, and the like. When presented with a post-sale photograph of an athletic shoe with four parallel stripes on the side and a shell toe, and asked the question, “Who makes or puts [this] out,” respondents do not meticulously review each feature of the stimulus but engage in “pattern matching” with respect to its salient characteristics. Respondents search their memory and identify the stimulus “based on its similarity to what [they] already know” and “[w]hen stimulus information offers a sufficient match to a schema possessed by the perceiver, the schema is called up from memory and used . . . to guide inferences.” As to the four stripes/shell toe example, 40% of test cell respondents typically draw the inference that the stimulus is adidas.

---

30 Jacoby, supra note 28, at 1035, 1037.
31 Hover & MacInnis, supra note 28, at 115.
33 Very little information can be held (and attended to) in active consciousness, referred to as “cognitive workspace” or “consciousness of the moment;” most information is stored in memory where its accessibility ranges from the instantaneous to the “virtually unavailable.” Jacoby, supra note 28, at 1015–16.
35 For a discussion and a diagram of the Nike brand schema, see J. Paul Peter & Jerry C. Olson, Consumer Behavior and Marketing Strategy 58–61, 74–81 (7th ed. 2005) (noting that a brand is a “bundle” of functional and psychosocial attributes).
36 Jacoby, supra note 28, at 1025; see Hover & MacInnis, supra note 28, at 183.
37 Peter & Olson, supra note 35, at 118–19.
38 For purposes of full disclosure, we have appeared as counsel for adidas A.G. in numerous cases.
39 Jacoby, supra note 28, at 1035, 1037.
40 Hover & MacInnis, supra note 28, at 102, 115–16 (3d ed. 2004) (“The cognitive networks in one’s memory . . . play a fundamental and often decisive role in interpreting incoming information from the outside world.”).
42 Expectation drives perception, Judith Lynne Zichowsky, The Psychology Behind Trademark Infringement 74 (2006), and respondents may overwrite features of a stimulus to conform to memory (e.g., they may convert four stripes to three). Swann, An Interdisciplinary Approach to Brand Strength, supra note 34, at 961–62.
3. **The Confusion Factors Tested by Eveready**

The Eveready format thus primarily addresses three confusion factors: similarity of marks, similarity of products, and brand strength (accessibility in memory). Strength is the key: (i) if a schema is easily accessible, it can be cued by a similar mark even where there is little or no similarity in products; and (ii) if a brand is dominant (COKE), its schema may be cued by another brand in the category (PEPSI), even where there is no similarity of marks. If, however, the senior mark is not accessible, it obviously cannot be cued irrespective of mark and product similarity: when an “open-end question [is] used [in connection with] a mark that is not particularly well-known, it needs to be understood that the . . . “top-of-mind” awareness of the brand . . . required [by the Eveready format] may significantly underestimate [the likelihood of] confusion.”

4. **The Scope of Eveready**

In cases involving strong marks, the Eveready format is the gold standard for fundamental cognitive and marketing reasons:

a. An Eveready survey (a) used among prospective consumers of the alleged infringer’s products or services; (b) in face-to-face interviews; and (c) with the stimulus left in view, engenders respondent “attention” approximating that of an “involved” consumer and thus produces, coupled with a control cell to filter market-share effects, a conservative (“reliable”) estimate of likelihood of confusion;

b. Reviews of “why” question answers typically reveal that senior mark responses to a “who makes or puts out” question have occurred because: (a) “stored knowledge” of the senior mark is “accessible” in a respondent’s memory; and (b) there is a “fit between the stored knowledge and the [junior] stimulus.” Accordingly, an expert’s conclusion as to

---

43 An Eveready survey can thus measure the “reach” of a strong mark. “A mark that is strong . . . is more likely to be remembered and more likely to be associated in the public mind with [or triggered by] a greater breadth of products . . . , than is a mark that is weak because relatively unknown . . . .” James Burrough Ltd., 540 F.2d at 276.


45 PHYLLIS J. WELTER, TRADEMARK SURVEYS § 24.03[1][c] (1999). Welter postulates that the Eveready format requires “unaided awareness” of the senior brand. In our view, however, the format more closely resembles a partially aided awareness test: it assesses whether the junior user’s mark and product cues are similar enough to those of the senior brand to trigger the latter’s schema in response to a source confusion question.

46 Jerre B. Swann, *Sophistication and the Sciences*, 97 TRADEMARK REP. 1309 (2007). It is our view that net confusion from an Eveready of less than 10% should suffice to support a conclusion as to likelihood of confusion (and that, because of an elevated degree of noise discussed below, a net level above 10% should be required from a Squirt format). See McCarthy, supra note 5, § 32.189 (discussing a 10% threshold).

47 See E. Tory Higgins, *Knowledge Activation: Accessiblity, Applicability, and Salience, in SOCIAL PSYCHOLOGY: HANDBOOK OF BASIC PRINCIPLES* 135 (Guilford Press 2007); Michel Tuan Pham & Gita V. Johar, *Contingent Processes of Source Identification*, 24 J. of CONSUMER RESEARCH 249, 250 (1997) (“The probability of source identification through cued retrieval depends essentially on [a.] the strength of the semantic link between the source and content that is formed at encoding, and [b.] the overlap between the cues that are available at retrieval, and the to-be-recollected material . . . .”). With ADIDAS, for example, most respondents give “stripes” in answer to a “why” question, reflecting their access to the three stripe mark and the “fit” or “overlap” with a stimulus that merely adds one stripe, or takes one stripe away.
a “likelihood of confusion,” based on an appreciable percentage of senior mark responses, has cognitively sound underpinnings;\footnote{48}

c. Because a strong mark is likely to be attended to in the marketplace,\footnote{49} it is reasonable to assume that a stimulus that “fits” the strong mark’s schema will be attended to, and that an Eveready survey thus measures probable assessments in the marketplace, not artificially created or forced opportunities; and

d. The only hypothetical is the degree to which a respondent would be likely to encounter the junior use in the marketplace, and any concern as to the real world basis for that likelihood is alleviated by limiting the universe to consumers and prospective consumers of goods in the category of the alleged infringer.\footnote{50}

With respect to strong marks, therefore, the Eveready format is a relevant, reliable, and objective test of likelihood of confusion. It satisfies critical \textit{Daubert} criteria as interpreted in the 2000 advisory committee’s notes to Federal Rule of Evidence 702: it is a “tested,” not a subjective, approach; it has been peer reviewed; with a control cell, it has a known error rate; and it has been generally accepted in the scientific community.\footnote{51}

\textbf{B. Squirt}

As befits the current conditions of marketplace clutter, almost two million marks are federally registered. Comparatively few have (or can hope to develop) sufficiently strong memory traces so as to be cued by pattern matching engendered by a monadic exposure to a similar junior use. The \textit{internal search} of memory for a strong brand’s schema that exists at the core of an Eveready study is thus hostile to the general run of marks; for weak marks, an Eveready format will consistently produce negligible estimates of likelihood of confusion. \textit{Ergo} the Squirt format, with an \textit{external review} of the marks at issue that flows from their side-by-side or sequential exposure inherent in the administration of a Squirt survey.\footnote{52}

\footnote{48} See, e.g., Jacoby, supra note 28, at 1028, 1034. A junior use may only trigger association (dilution), rather than identification (confusion), and questions as to what a stimulus “brings to mind” are appropriately rejected for likelihood of confusion purposes. \textit{Holiday Inns, Inc. v. Holiday Out in Am.}, 481 F.2d 445, 447 (5th Cir. 1973). “Confusion and dilution [and fair competition] . . . exist on a continuum,” and a junior user’s position on the continuum is predominantly a function of (a) senior brand strength, (b) mark similarity, and (c) product similarity. Swann, \textit{Dilation Redefined for the Year 2002}, supra note 44, at 620–21.

\footnote{49} See \textit{Swann}, \textit{An Interdisciplinary Approach to Brand Strength}, supra note 34, at 955. “[F]amiliar brands are selectively given more exposure, attention, comprehension and retention by consumers,” Steve Hoeflér & Kevin Lane Keller, \textit{The Marketing Advantages of Strong Brands}, 10 \textit{Brand Management} 421, 424 (2003), and owners of strong brands thus get “dramatically more impact from the same communications budget.” \textit{David A. Aaker, Managing Brand Equity: Capitalizing on the Value of a Brand Name} 186 (1991).

\footnote{50} Such a universe definition is thus a “relevancy” requirement. See, e.g., \textit{Zimmerman v. Nat’l Ass’n of Realtors}, No. 92032360, 2004 T.T.A.B. LEXIS 180 (T.T.A.B. Mar. 31, 2004). Likewise, where the junior user’s operations are geographically confined, the study should be confined to the area where there are respondents with the opportunity to come into contact with the junior mark. See, e.g., \textit{James Burrough Ltd.}, 540 F.2d at 277; \textit{Jacob Jacoby, Survey and Field Experimental Evidence, in The Psychology of Evidence and Trial Procedure} 181 (Saul M. Kassin & Lawrence S. Wrightsman eds., 1985).


1. The Questionnaire and Variants

In Squirtco, respondents first heard radio ads for SQUIRT and QUIRST 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?” In current designs, questions as to “sponsorship confusion” and “affiliation confusion” follow.

The number of Squirt variants is vast. In one, to remove a spotlight from the brands at issue, respondents are shown an array (including the senior and junior uses) and 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?

Another variant, to address objections to a side-by-side or sequential display of brands not seen in such immediate proximity in the marketplace, is a “two room” study: (a) in the first room, the respondent sees a stimulus of the allegedly infringed product; (b) in the second room, the respondent sees a “line-up” of products in the same category, including the allegedly infringing product, and is asked whether any “come from the same maker or company as the product . . . . I showed you [in the first room]? The two-room study “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 . . . .” Rather than rely on immediate juxtaposition of junior and senior
marks, it relies on a recent brand display (a “recency effect” in memory) to make the allegedly infringed brand accessible.

2. Categorization and the Representativeness Heuristic

As noted above, consumers do not meticulously examine brand stimuli; they categorize based on salient characteristics, and one mechanism they use in interpreting and classifying stimuli is the representativeness heuristic. A heuristic is a mental short-cut that consumers often take when making decisions, and according to the representativeness heuristic, consumers are likely to infer that things that are physically or conceptually similar or seem related must go or belong together. In *Beneficial Corp. v. Beneficial Capital Corp.*, the operation of the heuristic in a Squirt format was explained (in lay terms) as follows:

To the key question asked by the survey, “Do you think that there may or may not be a business connection between Beneficial Capital Corp. and the Beneficial Finance System Companies?” thirty-one percent of the respondents stated that such a connection was either definite or probable. The survey establishes that the names are similar and that portions of the general public will make the reasonable assumption that two companies with similar names are likely to have a business connection.

In *Wynn Oil Co. v. Thomas*, the heuristic arguably was enshrined into law: “[c]ases where a defendant uses an identical mark on competitive goods are ‘open and shut’ and do not involve protracted litigation to determine liability for trademark infringement.”

3. The Confusion Factors Tested by Squirt

A Squirt survey and variants of the Squirt format test similarity of marks, similarity of products, and market proximity. The proximity factor is critical. In an Eveready survey, given the “accessibility” of a strong mark, an unaided comparison (involving an internal search of memory) is appropriate where the respondent is likely to encounter the junior mark (and pattern match) in the natural flow of commerce. In a Squirt format, however, where the senior mark is not “accessible” in memory, an aided comparison (involving the representativeness heuristic) is appropriate where the marks exist side-by-side in the market or if one is typically encountered sufficiently soon after the other that the recent brand or stimulus exposure (the “recency effect”) places both in the consumer’s “cognitive workspace.” A Squirt survey is based on an external review of two stimuli that must be

---

62 See Hoyer & MacInnis, *supra* note 28, at 185 (“[Y]ou are more likely to remember what you ate for breakfast this morning than what you ate a week ago because (1) this morning’s information has not yet decayed [been forgotten], and (2) there is much less information interfering with the retrieval of this information.”).

63 Daniel Kahneman et al., *Judgment Under Uncertainty: Heuristics and Biases* 4 (1982) (“Many of the probabilistic questions with which people are concerned belong to one of the following types: . . . What is the probability that event A originates from process B? . . . In answering such questions, people typically rely on the representativeness heuristic, in which the probabilities are evaluated 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.”).


65 839 F.2d 1183, 1191 (6th Cir. 1988) (quoting McCarthy, *supra* note 5, § 23:3). But see Vincent N. Palladino, *Genericism Rationalized: Another View*, 90 TRADEMARK REP. 469, 478–79 (2000); Zeisel & Kaye, *supra* note 24, at 167–70 (negating the likelihood of confusion between “Workforce” for jeans sold in Gap and “Workforce” for socks sold in Sears). If the marks are weak and the products move through different channels of trade to different consumers, confusion may well be unlikely irrespective of mark and product identity.
substantially proximate for the review, under “marketplace conditions,” to occur. Absent market proximity, respondents in a Squirt design are made “artificially aware” of the competing marks.66

4. The Scope of Squirt

The historical distaste for “suggestive” questions67 is likely to continue with respect to such questions that have a clearly “leading” effect68 or to a Squirt without a control cell. With, however, the advent of experimental designs, the judicial hostility toward all closed-ended questions should abate. As noted above, such questions are often used as follow-ons to “Who makes or puts this out?” in Eveready designs, and closed-ended questions are typically used to test “comprehension” in surveys in false advertising cases.69

More recently, in National Distillers Products Co., LLC v. Refreshment Brands, Inc.,70 the court rejected a two-room Squirt study with respect to goods in the same category (vodka versus a vodka cooler) because, absent display of the senior mark in the first room, “respondent[s] would almost certainly have been unfamiliar with . . . [the allegedly infringed product] due to [its] very limited distribution network and weak sales.” That, however, is the reason for the existence of the format in the first place: without its insertion into consciousness, a weak mark cannot avail itself of consumer reaction evidence in a survey context.

Accordingly, many “closed-ended question” rejections and National Distillers are suspect law.71 The true limit on the design should derive from how the representativeness heuristic operates in a Squirt format—it facilitates inferences based on the similarity between the marks externally reviewed, either side-by-side or sequentially. A Squirt test should not be used, therefore, where the brands at issue do not proximately appear in the market. Under such circumstances, respondents who report a “connection” due to the “similarity of names” are “demonstrating merely that they had read the names . . . in artificially close proximity.”72

---

66 Kargo Global, 2007 U.S. Dist. LEXIS 57320, at *21–24 (“Kargo has offered no data or other evidence to support the proposition that prospective customers were likely to encounter Kargo’s trademark a short time after seeing Cargo magazine. . . . [I]t would [thus] have been far more replicative of actual market conditions to have displayed only Kargo’s materials and then asked the respondents open-ended questions regarding their beliefs about the source . . . . This is known as the ‘Eveready’ format.”). Because Kargo is weak, an Eveready test would likely produce zero evidence of likely confusion, but absent market proximity, only an Eveready reflects market reality. Where marks are weak and goods are not proximate, similarity assessments cannot occur either in the market or in the mind.

67 See Riviana Foods Inc. v. Societe Des Produits Nestle S.A., 33 U.S.P.Q.2d 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?”). There are only six instances in the last ten years where a court had relied, even in part, on a Squirt survey in reaching a conclusion that there was a likelihood of confusion.

68 See Proctor & Gamble v. Hoffmann-La Roche, No. 06 Civ. 0034(PAC), 2006 WL 2588002, at *22 (S.D.N.Y. 2006) (“The open-ended format is well suited for surveys focusing on simple and/or primary claims made in ads. On the other hand, open-ended questions are likely to understate secondary claims, particularly where those claims are rather complex by virtue of being both compound and comparative.”) (internal citation omitted). As Dr. Jacoby notes, “[r]eadily accessible stored information may be retrieved via open-ended (unaided recall) questions. However, retrieving less readily accessible stored information [as in a Squirt] generally requires using either ‘focused’ open-ended (aided recall) questions or closed-ended (recognition) questions.” Jacoby, supra note 28, at 1016 n.8.


70 Because the study in National Distillers had other flaws (the universe included potential purchasers of the allegedly infringed, not infringing product), the rejection of the study was correct. The court also noted that “the liquor and cooler markets are not coextensive” so there may have existed proximity concerns as well. Id. at 484.

A weak mark should not be placed in “recent” memory/“cognitive workspace” if it would not appear there under normal market conditions. To do so might be to give a weak mark artificial “reach” that it does not intrinsically possess. Quite simply, “in cases where the two parties’ 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.” Accordingly, the above-noted rationale that a two-room study “replicate[s] the marketplace process of [an] advertising exposure to a brand or trade dress, followed by being confronted in the market with both similar and differing brands or trade dresses” should be the subject of proof, not postulation.

Even where marks substantially and demonstrably overlap in the marketplace, we are still concerned as to whether they will be attended to, given that “the vast majority of stimuli fail to register upon the consumer’s consciousness.” Given the elevated “confusion” levels that Squirt studies produce, we are further troubled by a design that often reflects control cell “noise” of more than 25%. We nonetheless appreciate a two-pronged Squirt rationale: (a) for substantially overlapping marks, a Squirt has appreciable scientific underpinnings; and (b) it may be relevant to know, at minimum, whether two marks are sufficiently similar to one another so that, if they do “register,” they effectively will be considered the same.

II. Conclusion

Pattern matching (in the mind) and the representativeness heuristic (pattern matching in the external marketplace) suggest that for similar marks, particularly as applied to similar goods, some level of confusion is likely—consumers draw inferences from a “fit” between stimuli. The choice between Eveready and Squirt is dictated by where, as a matter of market reality, brand similarity assessments can take place.

The Eveready format is ideal for assessing whether, as to strong marks or as to marks that do not appear proximately or otherwise overlap in the marketplace, a likelihood of confusion is appreciable. As to such marks, an Eveready survey tests market reality—“what would be the confusion level if we were to allow introduction of the junior user’s mark given the current level of awareness of the senior user’s mark.”

73 Zeisel & Kaye, supra note 24, at 167.
74 See Leelanau Wine Cellars, Ltd. v. Black & Red, Inc., 452 F. Supp. 2d 772, 784 (W.D. Mich. 2006), aff’d, 502 F.3d 504 (6th Cir. 2007), rejecting a two stage Squirt because there was “no indication that in the actual marketplace, purchasers of Chateau de Leelanau wine are ever exposed to LWC’s advertising shortly before they view or purchase [the] wine.” In a 1989 case, Dr. Zeisel was confronted with the result from a side-by-side Squirt study that 35% of respondents believed that “Workforce” for blue jeans sold by Gap were from the same company as “Workforce” socks sold in Sears stores. He then conducted two studies “to measure the likelihood that consumers would ever encounter both products” or would be aware “of the use of the mark on both products;” found that the likelihood ranged between .5% and 1%; and concluded that “the likelihood of consumer confusion was not 35%, but less than 1% among both Gap customers and the general population.” Zeisel & Kaye, supra note 24, 167–69.
75 Jacoby, supra note 28, at 1034.
76 See, e.g., Pep Boys v. Goodyear, 2002 U.S. Dist. LEXIS 5925, at *30. We understand that there are natural error rates and subconscious and subliminal influences on memory, Hoyer & MacInnis, supra note 28, at 97–98, 433, and that elevated noise may be just a byproduct of general marketplace clutter. See, e.g., Jacob Jacoby et al., Am. Ass’n of Advertising Agencies, Miscomprehension of TeleviSeD Communications (1980) (“the average amount of miscomprehension associated with any [televised] communication was an unexpectedly high 30%”). “Reliability” and elevated control cell noise are, however, disconnects.
The Squirt format is the alternative for testing the likelihood of confusion between marks that are weak, and cannot thus be compared unaided in the mind, but that substantially overlap or are simultaneously or sequentially accessible in the marketplace for comparison—so that their aided presentation is realistic, not artificial. It relies on the “proximity” factors in a likelihood of confusion analysis, rather than on the strength factor, as its market replication rationale.

That leaves, of course, a question as to how likelihood of confusion should be tested as to a weak mark that does not appear in proximity to a similar junior use. We know of no format for that purpose and, as may be gleaned from the foregoing, we consider it appropriate that none seems to exist. Compliant with a market replication mandate, there simply appears to be no way to test whether a weak mark will be confused with another’s use in a commercial arena where the weak mark does not appear and there is no customer overlap. Trademark law has long considered the relevance of geographic demarcations on a macro scale.\(^\text{78}\) With respect to choosing the appropriate format for a likelihood of confusion study, it may sometimes be necessary to engage in a micro analysis.

\(^{78}\) Dawn Donut Co. v. Hart’s Food Stores, Inc., 267 F.2d 358 (2d Cir. 1959).