The Dimming of the Marketing Enlightenment of 1945-69: Sales-predictive tests and brand demand theory

Robert Hopkins, Allstate Research and Planning Center, USA

In the 25 years after World War 2, pre-tests of new product and advertising ideas advanced beyond asking consumers to state their preferences. Sales predictions were found to be accurate when physical response was measured (pointing or electrodermal response). Marketers later retreated to now-conventional interviewing, despite poor sales prediction validity and 80% likelihood of marketplace failure. The major question is: Why did marketers then (and more so now) prefer less valid verbal data over relevant non-verbal data? Two answers are proposed: 1. Managers and marketing educators depend on words for authority in workplace and class to such a degree in easy times that logocentrism prevails over better measurement. 2. Explanation for test results was needed but emerged only recently in psychological theories of framing, modularity and semi-consciousness. These proposals suggest a theory of free market demand that escapes from social science to find relative liberation within a frame of privacy.

A viewpoint. We gain important insight into marketing intelligence in the mid-20th Century by comparing marketers with 19th Century physicians. Both specialties experienced pre-modern phases. Medicine turned modern on the basis of science: empirical facts and coherent theory. In marketing -- the part concerned with new products, packages and messages, at least -- we are still mired in practices and theories that fail to meet the core challenge.

Pre-modern physicians were quacks. Most of us have heard the comment that surviving a doctor’s treatment over 200 years ago was in fact evidence of the body’s durability, not good medicine. This changed after many breakthroughs, most of all the germ theory revelation we credit to Pasteur, Fleming and others. Psychosomatics and natural healing work on some diseases, but getting rid of bad bugs was a key to advance ... quackery yielded to management of infection.

Recognizing bad bugs and finding ways to get rid of them are still needed in consumer research on new developments. Some people claim the field is already modern, such as contributors to the journal Marketing Science and the Marketing Science Institute (MSI). A case for such a claim should include evidence of general marketing success -- in the sense that marketers now reliably achieve profitability with new products and new advertising campaigns. The evidence we have is very different.

Current reports on new products or new ads (for old products) indicate that marketers typically don’t know how to overcome the natural probability of red ink, roughly 80%. Marketers need abundant intuition, talent, insight, experience, capital and organization but, in the marketplace, these resources are enough to profit only once in five conventional tries. CEOs should, but rarely do crack down when marketers can’t lift sales profitably.

Reform in medicine was part of The Enlightenment, my historical metaphor. The purpose is to find facts to improve on guesses and folklore. Schopenhauer viewed it as harsh reality defeating glorious ideals. C.P. Snow (1959) fretted about Two Cultures, science and the humanities in conflict.

Marketing also developed two cultures after the war. One was evidence-based, grounded on good data. Make-believe was the other culture. Make-believe proved more tenacious after 1969. The enlightenment is almost dark.

The development of sales-predictive consumer response testing moved marketing toward modern science in the late 1940s. While many demands of marketing are more expensive and complex, research must tell managers what consumers will be buying anew in a few months and if it will be profitable. The technical heart of the matter is: When new ideas arise, what, exactly, do we measure to forecast sales? Today’s answer is: many popular verbal response measures of attractiveness that do not predict sales well and a few highly predictive physical response measures that are unpopular.

Our pre-test goal is quantified sales appeal that addresses the profitability threshold. Without a measure that predicts sales, research falls short of helping marketers advance toward the fundamentals of science/technology, prediction and control. Predictive failures lead mystics, especially among ad agencies, to argue that it is enough to communicate conscious, verbalizable impressions. Apart from the prediction issue, however, history shows the control issue has been harder for marketers in terms of self-control.

Tests mostly collect consumer words about new ideas (wear-out tests are exceptions). Conventionally, the raw
data are the spoken or written words recorded in interviews and later interpreted by marketing managers as indicating sales likelihood. Some answers prove better (Haley and Baldinger 1991; Lodish, et al. 1995) than the incantations people tried for relief from infection before medicine’s modern era, but all verbal predictors fall short of perfect.

Among many kinds of management mistakes, relying on verbal expression from consumers is a big one. We can’t fix it with elegant statistics. GIGO leads to multiplying by zero. Without better measures, marketing can’t avoid nature’s apparent 80% probability of economic stumble.

This mistake can be corrected by measuring any of several physical gestures that are non-verbal expressions of interest or attraction. Older scholars dismissed the physical sales predictors they knew, mostly because informed marketers disregarded them in the 80s (Lipstein 1985; Singh and Cole 1988; Bagozzi 1991). Young marketers today are simply uninformed. More subtly pernicious causes for dismissal emerged from psychological research long after the heyday of non-verbal measurement, the 25 years after World War 2 ended.

**History.** The development of physical response tests largely resulted from the wartime crisis atmosphere. There was a hangover which induced a generation of marketers to continue avoiding waste after 1945. In addition, some producers and retailers worried that the depressed demand of the 30s would resume after military spending dropped.

Nervous marketers during those years knew they wanted accurate sales forecasts. Substantial progress resulted with a physical evaluation of radio shows just before the war when Lazarsfeld and Stanton (1941) used green and red buttons they called program analyzer. They could predict ratings.

Two other physical measurements were found to predict sales after the war (Eckstrand and Gilliland 1948; Cheskin and Ward 1948). Both new product designs and new advertising messages were tested, at first in sets of alternatives with results reported in terms of ordinal ranks.

Rank scaling is a modest level of measurement, compared with cardinal, equal interval and ratio scaling. At this humble level, physical data were consistently accurate while the best of conventional survey methods were not, then (Juster 1966; Axelrod 1968) and later (Haley and Baldinger 1991; Lodish et al. 1995). Despite then-available evidence of verbal method flaws, most marketers were not quite nervous enough and they relied on semantics.

George Gallup is a well-known researcher whose methods represent semantic measurement. Gallup started in the 30s at ad agency Young and Rubicam where he measured ad recall and favorable attitude change. The agency needed to give frustrated clients a substitute when ads failed to lift revenues (Ohmer 2000).

In addition to accuracy, physical response tests are both quicker and less costly than Gallup-style surveys in delivering go/no-go signals on new ideas before making huge production investments. Several were used widely enough to make their vendors wealthy: e.g., pointing with a pencil or gummed sticker (conscious motor gestures) was measured by Horace Schwerin, Louis Cheskin and Eric Marder; skin conductance fluctuation (a subconscious, autonomic response) was measured by Walt Wesley, Tom Turrichi and a few others.

Amazement at the atom bomb coaxed some marketers to trust methods that were viewed as black boxes. One was Wesley’s galvanometer, as known then. Marketers’ non-nonsense judgment on such measurement arose from a whatever-works pragmatism which persisted until the war and depression scares wore off.

Additional black boxes deserve mention. To test print ads in the 50s, Cheskin introduced eye tracking (1959; see also Treistman and Gregg 1979). In the 70s, voice pitch analysis (VOPAN) was added to test resources (Brickman 1976). Neither had much impact but the limited reports encourage further exploration. Speculation also led some people to measure pupil dilation, response latency, brain waves (EEG) and visual speed (tachistoscopy). Subsequent reports failed to show sales forecast validity.

Soft reports can be cited on new product/advertising failure during the 50s and 60s. Proctor & Gamble, a diligent pre-tester, was said to have reduced its new product failure rate to roughly 50% (Flesch, personal communication). That was encouraging. But more compelling were reports from witnesses who credited both Wesley and Cheskin with zero error in predicting rank-order sales results (Lindberg, personal communication; Hadley 1959). Evidence standards were not demanding then: They only meant zero so far as they knew.

Accumulating scores from Wesley’s tests over past years, we now see that the threshold of profit falls around the 80th percentile (Hopkins and Fletcher 1994) -- as it should, given good evidence (Abraham and Lodish 1992) that 80% of new ideas lose money. Conceivably, with accurate tests and discipline, management can avoid most, maybe all, marketing red ink. So why did marketers (and their teachers) reject the most accurate tools needed to efficiently market new ideas?

**Demand side research.** The content of verbal pre-tests is usually considered to consist of the semantics (dictionary definitions or face validity) of spoken/written brand awareness, advertising recall, attitudes and intentions to buy. In requiring semantics, managers get an office-like dialog which, in effect, corrupts the demand side standard of marketplace choice that gave substance to the marketing concept of the 50s. (The concept amounted to what is now called “behavioral economics” in Figure 1.) This retreat was facilitated by the historic failure to find the theory necessary to explain why good pre-tests work and bad tests don’t — explanation needed once military/economic crises and the thrill of the atom bomb wore off.

**Supply side research.** Hovering over supply in Figure 1, we see the nasty rate of roughly 80% failure to profit from marketing as usual. Red ink. (Figure 1 omits some
research tasks such as brand share tracking for old products—i.e., cash cows that pay for unprofitable new ideas—as well as classical supply and demand price gradients.

![Diagram](image)

**FIGURE 1. ECONOMIC FRAMEWORK FOR CONSUMER PRE-TESTS OF NEW MARKETING IDEAS**

After 1969, words steadily conquered consumer research. One notable example of how scholars exerted their preference was a 1986 advertising study by Stewart and Furse, sponsored by MSI. Only two theories, the hierarchy of conscious effects and cognitive response, were cited and all theories were dismissed for lack of specifying "the form of measurement". Strangely, ARS’s measurement was selected from 20 competing ad test services without recognition of how the ARS method differed from others. Sales predictiveness (external validity) was one criterion among several. The method was described only vaguely:

"Persuasion ... is obtained by asking respondents to indicate brand preference .... Products are presented in photographs that resemble store shelf placements. Both the product ... in the test commercial and competitive products ... are presented."

There was no acknowledgement of how physical expression (described below) differs from what readers would ordinarily read into the description, verbal declarations.

Since 1969, we see relevant theoretical progress outside marketing. The recent progress realized the neglected recommendations of such economists as Francis Edgeworth (1881), Simon Patten (1889) and Hazel Kyrk. (1923). Following Bentham, Edgeworth advocated a concept he called a "hedonometric" which measured feelings on a cardinal scale. Patten and Kyrk were early champions of drawing from psychology, sociology and philosophy to move marketing theory beyond the classical economist's fixation on prices (Mason 1998, 2000).

Dissatisfaction with cardinal scales led Paul Samuelson (1938) and others to reduce their measurement of consumer preference to ordinal rank scales. At the ordinal level, we can tell better from worse but nothing more. It persists in today’s discrete choice attempts to evaluate trait utilities in trade-offs among alternative product designs.

**Schwerin.** Shortly after Samuelson gave up on cardinal scaling, Horace Schwerin started work with Paul Lazarsfeld and Frank Stanton who measured response to radio content. (Hollonquist and Suchman 1944). One of many methods

**Keynesian** (money supply, interest, tax, subsidy, regulation, prosecution) — very social

**Classical, Neo-classical** (auction, contract negotiation) — largely social

**Behavioral** (choice of vendor, product substitutes & brands) — largely private

* Business-to-business, merchandising research

was the program analyzer by which each listener responded throughout a radio show by pressing one of two buttons. A green button in one hand signaled good moments; a red button in the other hand signaled bad. Later, buttons were replaced by a rotary dial or rheostat.

Apocryphal reports credited the program analyzer with good predictions of broadcast ratings. By so doing, it approximated cardinal scaling in which the overall percentages of favorable response consistently related to rating levels (percent of households listening). Emphasis at the time, however, was on finding strong and weak moments for editing.

After wartime service doing research on finding messages that persuaded soldiers to follow orders that needed more than ex officio command, Schwerin went into the advertising test business. His experiences taught him that asking people to say or write their preferences often failed to predict later choices. There was one inexpensive alternative to the program analyzer that he found to predict sales response to ads just as well. This alternative, a checklist, became his critical instrument.

Participants in each ad test gathered in a small movie theater in New York City and were asked first to check the brand they would prefer if they were the lucky winner of a year’s supply of the product category. Checklists included all major competitor brands. After the pre-exposure choices were collected, a new ad was screened. A second, identical checklist was passed out for a repeat choice from the brand list. The first measurement roughly indicated the advertised brand’s current
market share among product users; the second indicated the ad’s influence in terms of share increase when the pre-exposure share was subtracted.

Checklist results weren’t quite perfect sales forecasts, but much better than consumers’ stated opinions. Scherwin published his key findings (Scherwin and Newell 1981). There is more: He recently estimated that the book covered only 20% of all the information that his company reported in 15 years of research bulletins to clients.

Many articles have appeared in the past 20 years by Margaret Blair who succeeded Scherwin as head of the descendant firm, ARS, now located in Indiana and doing well. Dilution of semantics resulted when ARS replaced the brand checklist with the package photos mentioned in the quote from Stewart and Furse (1986). Now, participants make their choice by drawing a circle around the picture of their choice (allowing TV ad tests to get accurate data from both literate and the normally excluded market segment of functionally illiterate consumers — a large segment). Validity “in the 90% range” was reported in more-or-less stable markets; “the 60 to 70% range” if competition is turbulent (Blair and Kuse 2001).

The program analyzer didn’t go away after Scherwin dropped it. Several, including MSInteractive in Michigan and Richard Wirthein in Virginia, continue to do moment-to-moment tests with dials. Millward Brown tests commercials with a computer mouse instead of a dial. Reports on sales predictiveness have not improved on Scherwin/ARS.

Cheskin. A method comparable to Scherwin’s checklist was developed before 1945 in Chicago by Louis Cheskin who initially aspired to be an artist. Finding the public to be unresponsive to his artwork, Cheskin looked into interior decorating and studied color response. This led to his descriptor (largely adjective) check-off method. In 1945 he opened a testing service called The Color Research Institute. He soon published articles (Cheskin and Ward 1948) and books (Cheskin 1957, etc.) that richly documented his work, sold well and were often translated.

The adjective checklist was often administered to participants in a pre/post exposure schedule (like Scherwin’s) with a single idea shown between. Each checklist included 40, 50 or more descriptive terms that clearly connoted attractive or not. Cheskin counted the number of checks as his dependent variable. In a pre/post setting, a test ad or new product design was required to elicit more positive checks and fewer negative checks from a high percentage of the participants: 90% was mentioned in one report and said to be both rare and high enough to predict sales success.

The semantic differential was developed soon after by Charles Osgood (1952) at the University of Illinois. Osgood put adjectives into polar opposites with a 7-point scale between. He drastically reduced redundancy with factor analysis. The Osgood semantic differential has a large following. Though widely taught and used today in marketing, it fails to give very accurate forecasts of sales results. By discarding redundancy and imposing a magnitude scale, Osgood’s academic version of adjective checklists lost sensitivity to marketplace behavior.

A third good method using consumer pointing response was devised by Eric Marder who opened his New York City company in 1960. In testing a new product against competitors, Marder asked participants to allocate gummed stickers in a constant sum exercise. Ten stickers must be distributed according to relative preference. Marder at last published many interesting details in 1997.

Wesley. The pointing and grasping gesture is the common physical ingredient in the methods mentioned so far. Categorically different is the measure of skin electrical response that was most thoroughly refined in the late 40s by Walt Wesley in Chicago. Still called a galvanometer by journalists (or lie detector by fiction writers), the instrument measures, in current terms, electrodermal response (EDR). Wires are pressed on two fingers and an imperceptible electrical current is input to one finger while conductance to the other finger is measured. EDR arguably expresses attraction during exposure to a print or radio/TV ad.

Wesley published only two ephemeral documents but was open and informative when I undertook extensive reports that were published in 1987 and 1994. A 1995 report by LaBarbera and Tucciarone in the Journal of Advertising Research included a large collection of test scores which were compared with sales results (success or not — Figure 2). Wesley’s client for these tests was the Johnson & Johnson Baby Products Division.

![Gray = success, Black = failure](image)

**FIGURE 2. WESLEY RESULTS FOR A CLIENT WHO CAREFULLY MEASURED MARKETPLACE RESULTS (FROM LABARBERA AND TUCCIAARONE, 1995)**

Other researchers who measured EDR in the 60s include Tom Turcich who tested music for Columbia Records and Gerald Lukeman at ASI who tested movies and TV ads in a Hollywood theater. Later, Bruzzone Research in California and Inner Response in North Carolina have tested TV ads. In Germany, Werner...
Kroeber-Riel (1979) made it central to his Consumer Behavior Institute at the University of the Saarlands.

All these researchers trusted EDR data and told stories of frustration with clients. Progressively more marketing managers decided they were "comfortable" with data on brand awareness, ad recall, attitudes or purchase intention.

For most people, galvanometry is a single peak response, something Abraham Maslow might appreciate. Academic study of such issues as emotion is mostly limited to determining the bigger response to stimuli X vs. Z. Wesley found that such an application failed to give trustworthy forecasts of sales. A single peak misrepresents attractors although it serves well to indicate repellants (electric shock).

Radio research looked at response frequency during 60-minute programs. Wesley also saw that tests of 60-second commercials needed to reflect responses throughout the exposure plus the two seconds afterward. This led to dividing the time period into short periods and building an overall score with both frequency and magnitude components. When he added frequency to his scoring, Wesley in effect extended his test to include both Pavlovian and Skinnerian learning experiences. It works just as well with stimuli such as posters, envelope art or telephone book ads that get as few as 3 or 4 seconds of attention.

Behavior. The Skinnerian ingredient is fundamental to Wesley's success as a sales forecaster. Response frequency is the critical indicator in voluntary action learning. Recall that Pavlov's fame centered on a dog's glandular response - a gradual increase over many exposures, much like memorizing in the Ebbinghaus manner. Pavlov ran lots of animals through mazes, but it wasn't until Skinner modified Pavlov's apparatus as a lever in a box that muscular (motor) responses became a frequency, distinct from the effort exerted in each single response. What alerted Skinner to the contrast between glandular and motor response was his rarely appreciated finding that such tasks as reaching for brand X are effectively learned in a single trial rather than gradually. Half of his test animals learned the task on their first trial. Others required a second or third trial but all-or-none is the pattern in motor activity (Estes, 1964). This is surely related to Wesley's ability to measure shopping attraction during a single exposure to an ad.

Pavlov's and Skinner's reports predated the war. A few marketers considered them relevant (e.g., J. B. Watson). Cheskin later (1959) theorized about "sensation transfer" as a Pavlovian result of good advertising but he failed to appreciate his test's more Skinnerian nature. Operant behavior insights now serve best to explain the empirical success of physical response tests for marketing. (Relative absence of symbolic, or semantic, mediation of response is another ingredient of each test's level of validity. Best is so simple a pigeon can do it.)

So how do we explain the failure of marketers and their professors to get it then and since? Professors, after all, are responsible for appreciating and preserving excellence. Why are students now allowed to ignore Cheskin's once widely read books? Why aren't scholars conducting impartial replications and building on them? Lots of questions still need smart answers, although some good explanations emerged after 1969.

Basic scientists continue to expand our understanding of the physiological importance of EDR (Hirstein, Iversen and Ramachandran, 2001; Critchley, 2002). There is no weakness in the science. There is weakness in marketer priorities that need to be explained by social psychology.

McClelland. A major force behind disdain for EDR was revealed in David McClelland's reports (1975, 1985) on power-seeking as a personality force in corporate culture. The revealing clue to extremists on this tendency is obsessive thinking about winning status and authority. (Greed is also a clue if, as we say, money is power.)

McClelland had celebrated achievers in his better-known early research (McClelland et al., 1953) but after 20 years of further study, he concluded that careerists did not deserve celebration. Rather, problem-solvers qualified as true achievers and constituted the second personality type in his final 3-group classification. (McClelland would have called Scherwin, Cheskin, Wesley and Marder genuine achievers.)

The third personality is social nurturer, usually female, with priorities on the welfare of family and community. Smother love is an extreme that is as abusive as power-minded authoritarianism.

Power lust was found by Robert Deutsch in an observational study of fire chiefs, auctioneers and others. He reported, "They're all more concerned with problem structuring than problem solving" (Shalit, 1999). I.e., problems justify intervention, not fixes. Control reduces to mastery in the troublesome sense of dictatorship.

Expanding on McClelland's conclusion about what extremists brood about, I would argue that a do-it-my-way disposition leads both power- and nurture-minded men and women to be obsessed with words, the common tender of organization short of using police force: Consumers are expected by power-seeking marketers to give workplace-like answers; i.e., verbal commitments such as bosses get from employees. Nurturing parents and teachers want similar commitment from children in give-and-take situations.

We can certainly agree that words enable communication in polite society (with lots of cosmetic spin on self representation). Most consumers in a free market, however, act privately and more or less antisocially. Not all, of course, but most. Words are only sometimes needed from consumers to express personal choice. Current framing theory (Kahneman, Slovic and Tversky 1982) can be read

---

1 Comfort is an understatement. When EDR testing is suggested, marketers often look like a sprinter whose legs have been cut off.
to suggest that tests relying on words cause a respondent's mental frame to refocus on social purposes. In terms of role playing and other responses, social psychologists have said much about dialog throughout the 20th Century.

Anyone who works on questionnaires knows how picky people are about wording. Executives dictate final wording only because they have the most power in the room during research planning. Consumers, too, are picky in fitting their answers to circumstances. If verbal interrogation prompts most consumers to report attraction misleadingly, it's the listening corporate executives who eventually lose.

McClelland's method is as noteworthy as his conclusions. His results shared with one exceptional sales success in relying on a verbal measurement. Just after the war, auto makers anticipated returning to cars after years of concern with tanks and other off-road vehicles. The big 3 corporations surveyed consumers on design preferences. Ford and Chrysler asked, "What do you want?" They found consumers clamoring for practical cars, thrifty to operate and easy to maneuver. Ford and Chrysler decided on boxy, plain sedans with small engines.

General Motors chose instead to ask, "What does your neighbor want?" GM was told that the neighbors wanted streamlined, glitzy and fast cars. GM's designer, Harley Earl, went on to make his reputation (still cited in GM advertising) with snazzy gas guzzlers. By using both creative talent and predictive data on consumer dispositions, GM easily outsold Ford and Chrysler (who had only talent).

Both McClelland's test method and GM's were classified as "projective". Both pretend to engage a person's powers of observation of others. However, we can surmise that people are poor observers and their reports represent the only people they know well: themselves. They honestly report inadvertently because they do not realize that they themselves are on society's stage.

It appears that the projective method is now just as rarely used as non-verbal methods in marketing research. Superficially, it should be as appealing to managers as other verbal methods. But power-minded managers expect self-expression from consumers to be overt and binding, as they are used to getting from subordinates in the office where private agendas are anathema. Marketing professors march to a kindred drummer in the classroom, for the sake of both order and benevolent concern for student progress.

Privacy. It is charming to read that economics is the queen of the social sciences (Frey and Benz, in press). However privacy is the essence of the free consumer market, especially when grocery stores and other retailers are self-service. Social science monarchy has little relevance here. She can rule the supply-side's internal pecking order, B2B (especially ad agency-client transactions) and government functions. But consumer dispositions are liberated from social constraint. Enough to be remote from both Derrida's logocentrism and the reach of social scientists. I nominate consumers to be the democrats of the privacy sciences.

Differentiating private from social should be easy. There was a 1977 special issue of the Journal of Social Issues (JSI) on privacy. Relevant to the present topic, it included conclusions that "privacy serves to maximize freedom" and "even the most fervently held private opinions sometimes undergo an amazing transformation in the light of public scrutiny". Marketers could benefit from a Journal of Private Issues. Would be good, too.

There have been substantial divergences among studies that suggest how different the social (supply side) is from private (demand). Take Asch's and Milgram's well-known findings on social conformity. Sherif (1936) previously found that visual perception converged when people are in a group setting but not in solitude. Asch (1952) found that many people (most, not all) verbalize agreement with an apparent social consensus even when society is obviously wrong. Milgram later (1983) paid people to take commands that smacked of Hitler. In their motor responses, most subjects, not all, conformed to horrifying degrees.

In contrast, a recent test of consumer decisions by Ratner and Kahn (2002) found that a social frame prompted people to choose (via written response) more varieties of candy than a private frame did. This preference diversity superficially contradicts the Sherif/Asch perceptual and Milgram behavioral conformity. It demonstrates explicitly a social/private divergence only implied by the different preferences expressed verbally vs. physically.

Richard Crutchfield could have helped marketers in the 60s if his conformity findings had been more appreciated. Inspired by Asch, Crutchfield (1955) decided to conduct experiments more efficiently and he did something that echoed Skinner's reworking of Pavlov's method – he instrumented response measurement. By asking subjects to push buttons or levers in response to presentations of group consensus, Crutchfield found that taste preferences resisted conformity to authoritative opinion even though, as Asch found, "objective" perceptual judgments were still corrupted by such advice.

Crutchfield's findings had no lasting impact, despite an apparent contradiction with earlier reports by conformity researchers. Edward Thorndike for one, found many people knuckling under to authoritative evaluations of artwork. Thorndike asked his subjects to declare their decisions, rather than to point. He decided, on the basis of social dialog, that most people's tastes are easy to influence. Measuring physical gestures, Crutchfield saw otherwise. (This does not deny the possibility of influencing changes in taste with stronger appeals than those Crutchfield used.)

The 1977 JSI special issue also included a paper by Berscheid touching on what she called "unobtrusive" measures. These are often called anthropological methods today. Observation is disguised so a person doesn't realize
his/her actions are recorded. It should be clear that, except for the projective approach (mentioned by Berscheid), none of the good test methods cited here are unobtrusive. Participants know very well but don’t care that their responses are being measured when checklist or skin conductance equipment is used. Most people don’t intend deceit, but subtle social priorities take over when they chose words about themselves and their feelings.

Situational framing is now a prominent issue among economists for explaining a number of consumer preferences that fail to fit the rational, classical models. Daniel Kahneman’s 2002 Nobel Prize underlined the importance of irrational behavior in the marketplace. Many of us know his work in terms of decision research and heuristics such as prospect theory. Prospective thinking accounts for different choices when, say, a win is anticipated instead of a loss. Notice that, however important the work may be today, the early investigations (Kahneman, Slovic and Tversky 1982) mostly focused on verbal responses. Some behavior, particularly gambling, has been interpreted as consistent (Camerer 2000).

A cynic can argue that Kahneman simply revitalized gestalt (or field) theory, adding emphasis to the barriers between gestals. In rarely testing physical behavior, the implicit suggestion is to view gestals as though they extend seamlessly from words to deeds. In fact, barrier seems run between consumer words and private deeds.

Deeds are also framed by circumstance. Some marketplace frames are less conducive to free personal choice than others – teen-agers buy differently in the presence of older relatives vs. best friends. An important advantage of using private demand measures is their sensitivity to medium and long run sales, a month or more following test. Sooner or later, most consumers find themselves shopping either alone or with confidants who impose no social constraint on what predictive data indicate to be personal dispositions.

Some physical test methods that provide privacy are in practice flawed because they depend on sales-like choices in such short periods as a few minutes. The Copernicus test lab is one. It descends from a Yankelovich, Skelly and White laboratory in which a sample of product users is shown a new idea. After presentation, everyone exits the lab through a retail setting. They are asked to use part of their participation incentive to buy whatever they want from the display, including the test brand among competitors. In 1985, the company was telling prospective clients that results corresponded to later sales in the marketplace about 90% of the time. Good but again not perfect. The immediate situation is too confining and the test lacks redundancy (much like the Schwerin/ARS method).

Modularity. Frame-like discontinuities extend beyond market tests. General insights have been developed in combined psychology and physiology studies: Modularity is another label for the compartmentalization found to separate functions (speaking vs. pointing and skin conductance). It puts emphasis on brain structures examined with functional magnetic resonance imaging (fMRI) that localizes processes in a way that harkens back to phrenology. We can now see that brain puts seams between experiences: Seams that we normally fail to notice subjectively in much the way we miss the blind spot (scotoma) in our visual field.

Bruce Bridgeman (1999), at the University of California-Santa Cruz, found a startling example of separately modularized speaking and pointing when he experimented with the Rolof’s optical illusion. In this dark room experience, a person is briefly shown a rectangle enclosing a dot. Then a second, relocated, rectangle is shown with the dot in the same place. Asked if the dot has moved, people say yes. Asked to jab at the dot, people point at the true location. People are not aware of the response’s correctness and so we might call it robot-like correctness.

So far in his experiment, modular separation was evident. But, as we know, some impulses can be controlled in the thoughtful decision process that some scholars today call agency and the ancient Greeks called krasia (will). Bridgeman asked his subjects to hesitate 2 seconds before pointing at the Rolof’s and found people then point off in the same direction they say the dot moved.

Instruction to delay 2 seconds converts the exercise from a fairly impulsive expression into a more complex, unfamiliar task. Bridgeman also requested a one-second delay but subjects were unable to harness physical tendency to will power. They tried, but the results show that most of us (if not all) need more than a second’s deliberation to subordinate physical impulse to new mental task management.

I discussed this with Bridgeman briefly and, though he did not agree with my interpretation, delaying one’s physical response to a Rolof’s illusion is clearly more of a problem to solve. With practice, the illusion’s effect should become evident in a one-second or half-second delay situation. In this, we can see a kinship with language and verbal response. Composing a sentence in a new foreign language is a problem that takes more conscious effort and time to solve than verbalizing in one’s native language—an experience that talkative people (like present company) come to feel is no problem at all.

What Bridgeman saw in his data is visual modularity. Visual information has been found to follow two major pathways in the brain. One is shorter, traversing the brain stem and then a dorsal path (top of the brain). This path is involved in locating a stimulus and guiding motor action. The longer path courses ventrally (around the bottom of the brain) through the temporal lobe where it intersects with language areas of the brain. The ventral route served Bridgeman to explain the similarity between delayed motor response and verbal response to the illusion. 2

2 Other researchers (Dassonville and Bala 2002) emphasized the verbal response linkage to memory.
Bridgeman did not integrate his findings into the procedural vs. declarative modularity that emerged over 20 years ago (Cohen and Squire 1980). However, Ten Berge and Van Heeswijk (1999) have done so within a discussion of the evolutionary functionality that gave rise to structural separation of motor behavior from conscious verbalization.

Modularity in childhood imagination was suggested by findings of Oxford’s Paul L. Harris (2000), now at Harvard. His conclusions apply to collecting test data from consumers because he decided that fantasizing emerges at the same time that kids learn to compose sentences. Composing a new thought draws on one’s creative powers. Whatever is perceived to be true, verbal extrapolation packages the truthful detail in a synthesized context with more or less distortion. We forget this about learning our native tongue but it is clear again when we study a foreign language and must compose sentences with new lexicon and grammar. It is harder to find words for truth than to choose a brand, especially if you are not talkative.

Extensive experimental findings of modular separation of perception from imagination in the brain have been reported (Gauthier et al. 2002). Also noteworthy are investigations by Blakemore and Decety (2001) who found brain loci for our penchant for imagining intent when seeing mere happenstance. A projective theory, simulation, was cited to suggest, as McCloud did, that people tend to constantly think about their own desires and imagine that such yearnings are behind everything that happens.


Modules routinely collaborate in distributed functions. Boston psychologist Drew Weston (2002) has suggested that this moved us to the second wave in the cognitive revolution. It is a partial behaviorism restoration. The second wave is far more useful to marketers than competing proclamations, such as the claim that all-talk, all-the-time (i.e., the narrative turn) is a second cognitive revolution, with social construction at its heart (Moghaddan and Harre 1992). What marketers need from this is to realize that consumers use words as social tools -- bad bugs to avoid in a consumer science, regardless of their importance for social pursuits.

**Semi-consciousness.** A kindred argument was recently offered to marketers in the *Journal of Consumer Research* by Bargh (2002) who focused on the level of consciousness required in such decisions as brand choice. Bargh has demonstrated that what he calls automatic behavior accounts for most decisions of the sort we see in the consumer marketplace. I challenged Bargh on this, only because automatic suggests that brand choice could be like breathing, a function that people can do as well asleep as awake. The suggestion is worse when the notion of unconscious dispositions is used to discuss what was called low involvement shopping by Olshavsky and Granbois (1979) and peripheral processing by Cacciopo and Petty (1985). One might as well say comatose.

Bargh agreed that semi-consciousness might be more appropriate than unconsciousness. He also accepts semi-automatic and conscious automatic. Another stab was intenational blindness (IB), coined recently by Mack and Rock (1999) upon finding many people, not all, failing to remember seeing such surprising things as a gorilla costume appearing in the midst of a basketball game. (Neisser and Becklen 1975 reported the first of such findings.)

Bridgeman found physical responses to be correct and fully conscious responses to be wrong. IB is a case of partially conscious response error. Right or wrong, there are meaningful differences in full vs. partial consciousness.

We are trying to nail down an in-between consciousness that is popularly called absent-mindedness. It likewise does not mean true absence. Robot-like conduct is a level of wakeful function with attenuated consciousness allowable while events seem to go routinely (with tolerable errors). Right or wrong is not a consumer issue, of course. Attractive or not is. We can accept semi-consciousness for private choice but not for dialog at work, in class or while answering market research questions. Normal social performance gets full consciousness. This conclusion contradicts sociologist Gabriel Tarde who said “social man is a somnambulist” (quoted in Asch 1955) in the early 20th Century when conformity appeared to be due to hypnosis.

Once we recognize that some pursuits can be semi-conscious, we can enrich the picture of mind and conduct with another finding that has also been hard to find good words for. In this case, where behavioral research has suggested separation, seams are absent or, at most, minor.

Again Daniel Kahneman helped: His examination of the integration of perception, learning and action led to a concept he called object files. In a study on experience that resists fragmentation, Kahneman and Gibbs (1992) concluded that stimuli and responses are meaningfully integrated: An attractive stimulus elicits approach and, if appropriate, acquisition in a way that becomes an organic component of the stimulus experience ... often, not always.

Another approach to large patterns of experience like responding to new ads was offered by Hommel, et al. (2001), who called it event coding. Citing object files as a precedent, these investigators added to the evidence that response from one’s behavioral repertory is integrated into the mental registration of a stimulus. From such studies, we can conclude that our experience of stimuli is often less remote from response than suggested by such computational cognitivists as Fishbein and Ajzen (1975) and production system modelers (Newell and Simon 1972).

**Quantum process.** An even larger integration to consider is William Baum’s argument, “From molecular to molar” (2002). Baum made this useful point: Operant or
motor responses can be equated to light, a combination of both material (or particle) and energy. As such, behavior proceeds through time and space in a larger continuum than suggested in the usual notation for stimulus and response (S-R, R-S, S-R-S).

Baum suggested we imagine a pigeon in a Skinner box (or a shopper looking in the phone book). The animal is likely to peck at something and its head (or shopper finger) rises and falls in a wave form much like a recording of electrodermal change. We can also consider music, games and careers to be repetitive process, like shopping. People pursue them with continuity that is not represented well in discussions that dwell on isolated episodes. The usual survey questions about brand sales appeal confuse response to a narrower frame than market experience constitutes.

Baum made it easy to reach out to physics for its lesson on quantum processes. Quantum theory tells us that individual consumer behavior is less than certain and we should seek wisdom in quantitative data, not anecdotes. This should discourage us from spending so much time on case studies as business students do today. For one thing, anecdotal reasoning is hind-sighted and cluttered with irrelevant impressions (as most questionnaires are). What we need are fore-sighted data that tell us our probability of success with a given new idea for attracting customers.

Probabilistic brand demand among product users can be stated as: \( B = EDR_w \). Economists would call this a measure of utility. Big ad budgets produce more sales than small budgets only when \( EDR_w \) is high.

The success threshold is based on the distribution of scores in the natural marketplace. This formula says brand sales correspond to electrodermal response as in Wesley’s measure (Figure 2). The good forecast methods as a class, F, share the functions of two modular frames addressed here, privacy (P) and semi-consciousness (S-C) which both preclude a wordy power frame. So: \( B = F(P, S-C) \).

Other demand functions, some social, must be added (e.g., affordability, legality, etc.) to be absolutely complete.

Formulas are abstractions that most practicing marketers dismiss as too academic to use in new product and advertising decisions. A better utility also emerged during the 25 years I am describing as a time of enlightenment. It handles probability as well as it supports self-control.

Practical probability. In 1967, Irwin Gross completed his dissertation (Gross 1967a, 1974) on the statistical implications of the 80% likelihood of advertising failure. He developed a multiple draft tactic to achieve profitable sales from every new ad campaign.

Gross concluded that seven independent creative approaches were needed to lift the probability of success to virtual certainty. For advertising, seven separate creative teams need to be assigned the challenge to develop a new message for a given brand. A later recalculation changed the number to eight (Ehrman and Mische 1989). All of the alternatives must then be given proper tests with representative samples of the intended consumers. No management tweaking of the drafts is allowed. The consumer decides, in effect, which creative idea is a winner. Managers do not decide. It is determined just as in the ultimate marketplace.

Gross (1967b) presented his findings to the Advertising Research Foundation. He was rebuffed. Ad people were insulted to be told that their creative judgment reduced to throwing dice on a crap table. Gross couldn’t span the culture gap between his own and ad agency pretense.

The 30s version of two cultures involved Paul Lazarsfeld when he put the program analyzer to work for broadcasters as a scientific guide for managers to attract bigger audiences and lift revenues. A close colleague of Lazarsfeld, Theodor Adorno (guru of negative dialectic and critical studies) opposed such content management fearing that bureaucrats must surely compromise quality (elite arts such as the classical avant garde). If Lazarsfeld had linked his media work to the conformity findings of Asch and Crutchfield in the 50s, Adorno should have then been placated. Conformism afflicts many, even majorities, but not everyone. Even in social settings, non-conformists persist and sustain exceptional ideas. Moreover, physical measurement gets past social conformism to reveal the unvarnished personal preferences that shape the market.

Gross showed demand for creative talent to be 7- or 8-fold when marketing is efficiently consumer centered. This is considerably more than what is, in effect, today’s 5-fold requirement of logocentric managers who can’t avoid 4 failures out of every 5 tries. Artists will be as unhappy to see 7 ideas fail as they feel when 4 go nowhere, but lifting marketing efficiency should improve their incomes.

Humanist marketers who feel uncomfortable with science and technology should be offered this insight from The Enlightenment’s Francis Bacon: “Nature to be commanded must be obeyed”. Unlike Kant who thought authority was escape from nature, Bacon realized that science is a tool for management that provides good evidence but also requires self-control.

Even medicine continues to find practitioners making impulsive, impressionistic mistakes of the kinds explored by the decision researchers. Anecdotal thinking is a large part of the problem. To guide doctors toward better self-control, a movement called evidence-based medicine has emerged, mostly in England so far. The objective of movement leaders (McGovern et al. 2001) is to make statistical data on diseases and therapies the routine guide to prescription.

Conclusion. I suggest that if medicine were no further toward competence than marketing is, we would today be haunted by a much higher infant mortality rate than we are. Less than 80%, perhaps, but horribly high.

Predictive testing started to enlighten marketing with the demand side data needed to select winning new ideas from a bunch of mostly losers. Distractions included the supply side need for language and widespread appetite for social power. Lucky marketers survived because their firms enjoyed at least one cash cow. Surviving marketers...
settled on measuring consumer preferences in meeting-like dialogs or written interrogation.

Multiple drafts offered to organize the creative and test elements of new idea marketing so that every project succeeded. By taking the decision out of the reach of managers before huge production expenditures are made, this tactic was ruled out as an insult to expertise and a threat to executive prerogatives.

Together, predictive measures and multiple drafts offered, and still do, to help marketers muster the self-control required by new idea success in the democratic consumer market. With them, marketers can gain something analogous to a targeted contraceptive that prevents the 4 out of 5 product and advertising conceptions that are doomed by nature to be still born.

REFERENCES


Blair, Margaret H., and A. R. Kuse. 2001. Better practices in advertising can change the cost of doing business to wise investments in the business. Evansville: ARS.


The Romance of Marketing History


