ONLINE RESEARCH … AND ALL THAT JAZZ!

THE PRACTICAL ADAPTATION OF OLD TUNES TO MAKE NEW MUSIC

Steven Gittelman
Elaine Trimarchi

PREFACE

It is difficult, if not impossible, to determine the representative qualities of online data. The probabilistic models that once underpinned our research are re-visited. Here we take a hard look at the use of behavioral profiling of respondents to augment the old and create new representative sampling frames. We are at a crossroad; either we ground our research with proper theory or we lose to those who claim to read the wisdom of crowds.

INTRODUCTION

We are in the midst of a social revolution. The online universe is changing our world. For us in market research, many are at a state of dissonance, unresolved and uncomfortable; the probabilistic sampling frames of the past have eroded. They were our structure. Where once we could rely on a theoretical underpinning to give our work a solid standard, we now float in an unstructured and anxious space where we cannot gain our footing. We question the quality of the samples and are seeking ways to reassure ourselves.

We are at the fulcrum. We can either leverage what we know about structured, representative, market sampling or pit ourselves against the random noise of the crowd. Our advantage in developing a structured approach to sampling is massive. We understand the essence of a good sample frame and should know how to apply it. Working from a history of knowledge, we can provide order to the music of the crowd.

Have you ever listened to Gershwin's Rhapsody in Blue? That clarinet in the beginning is dissonant. There is structure in the music but also a vitality and newness of Jazz. Gershwin blends the old formal structures of music with the new world.

We have to follow Gershwin's model: take from the old and blend with the new. In the world of marketing research there will certainly be detractors but we must move fast as the net will not wait. We need a plan. In the jargon of market research, we need a sampling plan. As fundamental as such a concept may be, we need to agree as an industry on what this plan should be. Can we accept unstructured data emanating from a social network site or is there a need to understand what that data represents? There can be no more fundamental question in market research.

This is the fork in the road. Either market researchers of the future will be able to describe the representative nature of the data that they collect, or they will lose a tremendous repertoire of interpretive tools that tell us what the data means. In short, the ability to make decisions from that data will be impaired and the value of the profession could be diminished. Those who pursue the interpretation of unstructured data that cannot be referenced to a standard will have to tolerate greater risks in their decisions. In essence, market research always has and will continue to be in the business of risk management. We sell better grounded deductions based on the science of interpretation.

THE HISTORY OF SAMPLING

But where do we begin? Let's look back to those earlier sampling methods that we replaced; those better grounded in theory. We need to understand the importance of that theory and reconstruct new approaches that simulate the benefits inherent in the methods we have discarded. The most proximal shift was from telephone interviewing to online panel research.

Those of us who practiced in the era of random digit dialed telephone samples know that there was comfort in not having to defend the theory behind the methods used. We sought to reach each member of the sampling frame with equal probability. In the United States, we could measure our success against a census that was controlled by Constitutional Mandate. It cost billions and seemed unassailable. In theory, if we made a substantial effort to reach every respondent at their home, some 98% of them could be accessed through the proliferation of land lines. However, if we called at our convenience rather than theirs, we were more likely to reach those who were most accessible and create a non-response bias. As refusal rates climbed during the 1990s, the prospect of achieving the sought after recovery rate that implied equal access to all respondents began to erode. There were methods that we tried as we sharpened our tools to combat the challenges before us, such as call backs at different times and converting refusals with special teams of interviewers. It all seemed so difficult and yet in retrospect it was all so easy.

Online research is not the root cause of the erosion of our probabilistic framework. Its origins go far deeper. Alas, telephone ran into its own troubles with increasing refusal rates, do not call lists, technological methods of blocking calls, caller ID that separated friend from foe and now the new generation of problems caused by wire cutters who insist on using only cell phones to communicate with their world. Costs to do a well nested telephone study have risen so high that they are becoming a luxury affordable in only critical situations.

Online research has been troubled from its advent. Nothing that we can wish will provide it a true probabilistic frame: Those who refuse to go online provide a non-response bias akin to the problems that telephone now experiences with those who range free with their cell lines. It made sense in the beginning of online research to ground our comfort zone in comparisons between phone and online. There was a sense of credibility achieved when we could tie ourselves to an earlier trusted method. We were misguided, however, as the telephone standard had evolved into no more than a reference replete with its own problems. It is still a good thought to look for corroboration between modes but no mode is a true standard.

It appears that a combination of modes is required to approach our goal. The Census in the United States is a multimode study. As grand as it is, and as nearly flawless as it hopes to be, it only measures demography. It does not attempt to capture behavior, at least not to the degree that we as behavioral scientists are compelled to do when we complete a market research study.

The difference begs the question: Why do we rely completely on demography to create sampling frames when we are attempting to measure behavior? Should we pursue a behavioral standard? If all respondents of the same demographic cell were behavioral equivalents, then market research would soon lose its raison d'être and we would likely be earning our livings by a different manner. We measure behavior and it is the understanding of the behavioral differences and the variables that influence those differences that are at our core. Variability in our samples and bias in the way they represent demography and behavior within those populations, both contribute to noise that misleads us. We must understand that noise and bring it under control. We must create a viable standard where demography alone no longer serves as our only metric.

Online sample frames are fluid. Recent data from the Advertising Research Foundation made it clear that the online panels are not interchangeable. Sample sources seem to have a never ending repertoire of drivers of variability. The list continues to grow: respondent tenure, survey taking activity, the sources from which respondents are drawn, the manner in which they are recruited, the interest that respondents have in the subject matter at hand, engagement levels, incentive structures and on and on. We can slave to understand each of these in turn but unless we come up with a common language and consistent metrics, we won't be making progress for a while.

The new market researcher must learn to know his respondents as well as rebuild the theory that supports the frame. Discussions of quality are incomplete without an understanding of the sample frame and how it resonates within the theory that underpins our research. We are operating without a "safety net" and must create a sound foundation. What could that safety net be? Think of it this way; if we emulate the census we have a model as to how things should be. If, through a random selection of respondents where each respondent has an equal opportunity of selection we achieve a close simulation of the census, then there is the likelihood that we have achieved an acceptable representative sample frame. The census, where it was available, has been our "safety net" or standard to which we reference our sample distribution. If our sample is different from the standard, then the difficult questions begin.

Historically, one measure of the precision of our probabilistic sample was that we maintained counts of demographic cells from a census. In time, that transitioned into quotas that were a poor substitute for recovering random samples of the population and replicating our census. When we transitioned to non-probabilistic samples, we substituted adherence to high recovery and equal probability of representation with "completion" of quotas.

Quotas are often set in order to ensure that the sample is a match to a model, most often a census. Although all forms of market research data collection (including phone, mail, online and in person) produce a non-response bias, quota sampling theoretically provides the researcher with the ability to ensure that he has data about the population in which he intends to speak.

Quotas are a statistically controversial method, given the intrinsically non-random nature of the data collected. For example, the data collected using quotas will be devoid of respondents who were rejected because their respective cell was already filled. Given that respondents are included on a "first come, first served" basis, this means that the data used is not a random sample of those who could be reached, but rather those individuals whom the researcher was able to reach most easily. This may ultimately bias any inferences one wishes to make from the data if there is a difference between those who are included and those who are not.

Quota theory is familiar ground, though many use it without a comprehensive understanding. Can we build a viable sampling frame from it? Model based sampling that relies upon quotas existed long before probabilistic structures made life so easy. For those of us whose careers span only the online revolution, the concept of a sampling frame seems a luxury. Putting it aside, despite the beat of academic drums, is quite tempting and can easily be done.

The transition to online sampling was overseen by an eclectic assemblage of practitioners. Many argued that validating online research by comparing it to a phone study is problematic. The choice of variables used could not cover all possible situations. The possibility existed that the comparisons were among hyper-stable variables that were not subject to the biases of the new methodology. Those in academics scoffed at the method entirely pointing to obvious areas of bias. Today's new generation of market research practitioners grew in an age where sample is a commodity and the sampling frame was of diminishing importance. They were not exposed to the decades of probabilistic sampling practice that helped us achieve our goals with a theoretic underpinning that was intuitive and executable. Instead, we have threatened our own standing as professionals by failing to recreate a best in class sampling frame that withstands many of the challenges before us. For the most part, the new generation of market researcher does not understand the issues at hand, or worse, fails to value them. Many in the industry continue to fail to acknowledge the non-probabilistic nature of online samples and for some of those who have, the epiphany has occurred only in the past few years.

NON-PROBABILISTIC SAMPLING FRAMES

Non-probabilistic sample frames do not access all respondents with equal probability: it's all in the name. We have crudely substituted demographic quota control as a means of achieving a simulation of the census which is the only nearly probabilistic model we know. We have reluctantly retreated as price and timing pressures trumped theory.

It is that ease that lulled us into the place that we now find ourselves. Had we been forced to understand the rigor of non-probabilistic methods, we would have created a diverse framework of learning what would now serve us well. Fortunately, the rest of science lived in a real world where non-probabilistic, hypothesis based models rule. They now represent a rich treasure trove for us to explore. Sadly, we find ourselves lacking an understanding of how probabilistic methods function. Yet, we must move forward with non-probabilistic methods which are likely to be more complex than those required by our earlier underlying theory.

Perhaps a good fish story is relevant at this time. Population ecologists are trained to sample without the aid of a census. Fisheries biologists put out a large net to capture fish, tag and release them, wait, do a second netting and through the proportion of tagged fishes captured against the total in the net, estimate the population. Again, some fishes fall prey to the net more easily than others. The capture and release process is a random sampling of a biased subset. It is not a census. Other methods provide a better census of fish populations: draining, dynamite, poisoning and electric shock. Even these draconian methods have hidden bias, as they still retain an element of bias in the recovery process: little fishes may not be found as easily as large ones.

The capture and release method can be tested. When ponds are drained, a program of capture-release-recapture is employed, analyzed and compared to the more rigorous counts obtained when draining the pond. The point being that the methods themselves must be compared to a reference (draining) to calibrate the inherent bias. This re-calibration process is essential to understanding the applicability of the sampling frame. It becomes the language of interpretation of the data.

In that framework, let's take on a few challenges. There is an onslaught of respondents coming from well-meaning sources: social network, opt-in panel, river, phone, mail, and in person. How can we make these disparate groups talk to one another through a common language? If we fail to achieve the dialogue, then we will suffer the equivalent of a market research "Tower of Babel". We must become the tour guides through that tower. If we cannot master its language, then we can expect to be commoditized. We will surrender the reliance that business has historically had in us to be their eyes, the trust of the pathfinder. In our case, we must achieve an understanding of our sampling frame. In this fundamental re-invention of ourselves, we must find common denominators between sample sources that transcend the differences of the sourcing themselves.

Where once non-response bias was "the tree falling in the woods", we must now make a science of understanding how it affects our data. That is, since we did not look for it, we felt it could be ignored. How and whom the panels provide us within our sampling frame is no longer a matter of trust; it is a passion of our art. To do our jobs correctly, we must know and understand our respondents and of equal importance, know how to combine them.

WE NEED A STANDARD OR REFERENCE THAT SERVES OUR PURPOSE

Is a census the perfect standard? Perhaps we should accept the census as a standard and couple it with a standard of our own design; one that works better for all of us. Certainly, the census is designed to provide a language of demography that helps us when we speak about behavior. If, as described by the Advertising Research Foundation (Rubinson et al, 2010) "the panels are not interchangeable," differences that exist in the behaviors that we find within demographic cells can be source related. This drives us to a necessary change in paradigm. We must take a lesson from our past and create a new, workable, set of sampling frame standards.

Our industry is accustomed to categorizing behavior through structural segmentations. Thus, the combination of demographically balanced samples from different modes that provide us with a census of behavioral segments could assist us in inventing the standard that we seek. We need to approach samples that will transcend the sources of variability inherent in the sample sources themselves and become demonstrably representative. The word approach is intentionally chosen, as there never has been and never will be a perfect standard. Our ability to achieve perfection was more a matter of art rather than science. There has always been some demographic group that eluded our efforts. The less the effort (or the more poorly guided it was) the shorter we fell.

This brings up an interesting point of concession. The perfect probability sample has become a theoretical dream. Instead, we live in a world of "fit for purpose." In essence, to move forward, we must accept a bit of dissonance in our thinking. We need to be willing to leave some things unresolved. Our standards will have tremendous pressure applied to them right from their very creation. The United States Census is not replicated throughout the world and yet we need a global standard. The end users, who purchase our research, are most often global in scope and will seek standards of global reach. There is no one-size-fits-all approach and thus we argue for latitude in our creation. In fact, we need to evolve as the fast changing world continues to rotate.

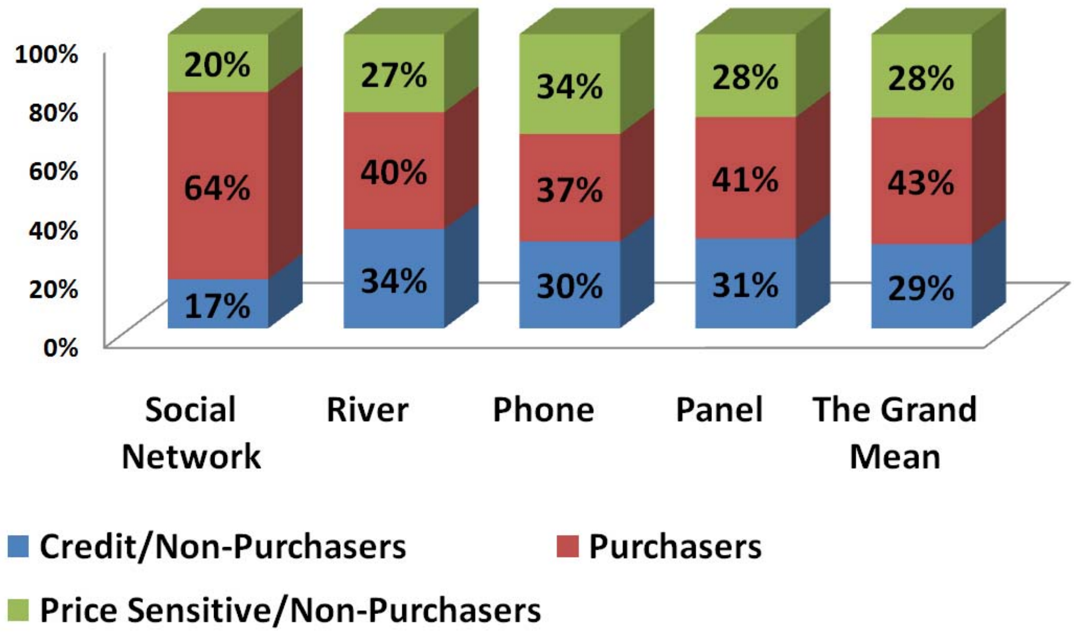
To do anything else might leave us woefully behind in what has become an intellectual race to master the data. If we continue to seek perfection, then we will fall short on execution. In a sense, just as dissonance became an important thematic component in jazz, traditional music experimented with dissonance and was influenced by it. The challenge for market research is to adopt past theory to the new music. The current weakness is that we can no longer measure behavior through a demographic standard.

We propose that respondents should be behaviorally pre-profiled and classified by a battery of segmentations that we distribute according to a demographic/behavioral standard. To this effort, we use a battery of ten segmentations in our scheme that include three general segments: buying behavior (37 variables), sociographics (31 variables), media (31 variables), and seven market segments: automotive, appliance, consumer electronics, clothing, grocery, entertainment, insurance/banking. (When required, such as in social network and river sampling methods, a real time abbreviated segmentation scheme can be employed. However, this comes at a cost because the fewer the input variables, the less stable the segmentation.)

We use a sixteen minute questionnaire designed to generate the ten segmentations geared to be sensitive measures of sample source change. The questionnaire has been translated into the languages needed to execute in 35 countries and we have obtained the cooperation of some 200 panels around the world. As an example, let's focus on survey research in the United States. First we have to create a standard.

We employ telephone and online sources as our modes of choice. Online sample was further divided into three sample segments: river, social network and opt-in panel. The four sources are behaviorally different (figure 1) and have been combined in the following proportions: Opt–in panel, 40%; Phone (includes landline and cell phone), 30%; River, 15%; and Social Network, 15%. As this is the combination of sources, we have named it "The Grand Mean."

FIGURE 1
BUYER BEHAVIOR BY SAMPLE SOURCE



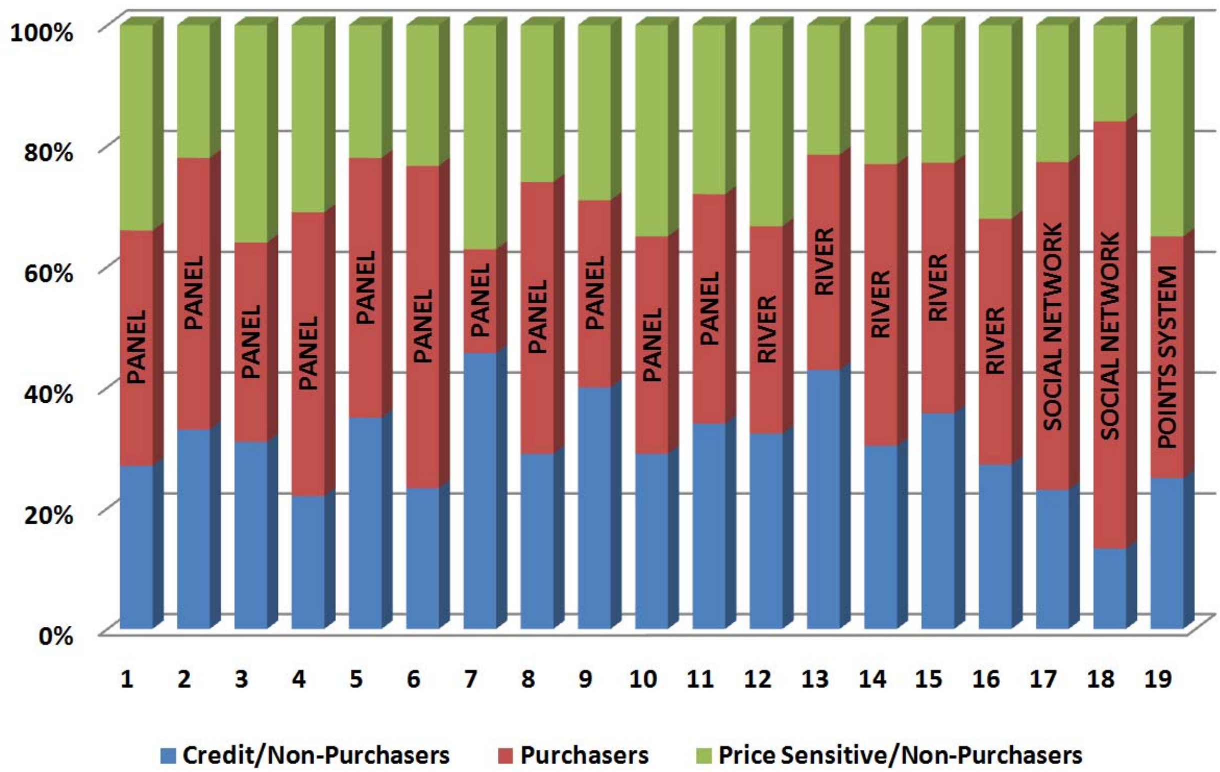
Distribution of the Buyer Behavior segments by sample reference and the Grand Mean.

The segments become our common language and the distribution of the segments is a form of census. The demography provides us with cells, each having a distribution of segments which we construe as being representative of the population in question.

It would be quite useful to create the Grand Mean in combination with mail and in-person data, in those countries where such data can be collected. In the United States, a rigorously applied telephone study combined with an online sample, where both have highly nested quota controls, has been selected as a reference.

We can now assess the differences between sample sources by comparing the distribution of their segments. (See figure 2.)

FIGURE 2
BUYER BEHAVIOR BY RECRUITMENT METHOD



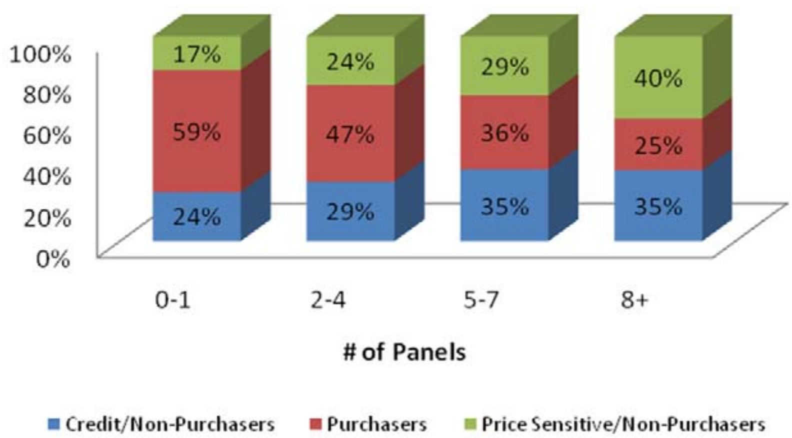
Individual American sample sources (phone, individual panels, rivers and social networks) compared by differences between buying behavior segments.

We can measure drivers of variability by the changes in the segments that they create.

As an example in figure 3, buying behavior is shown to shift by different measures of respondent activity. Those who participate in more surveys or belong to more panels exhibit different buying behavior segments from less active participants. This difference occurs across sample sources (social network, opt-in panel).

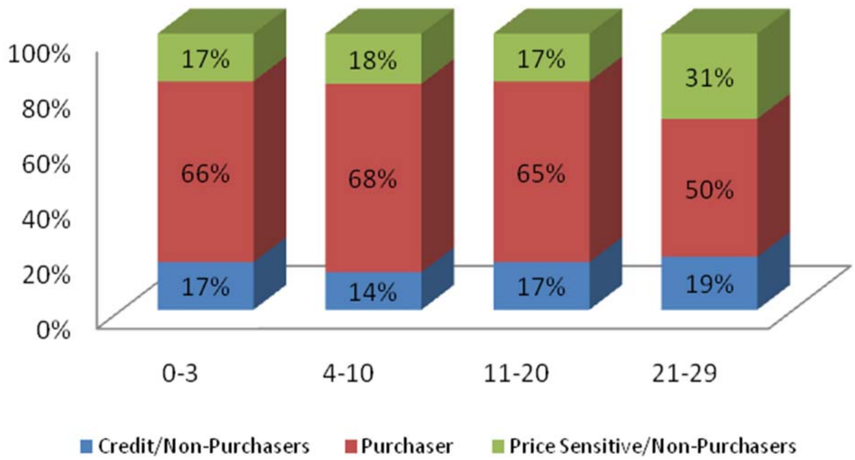
FIGURE 3
BUYER BEHAVIOR BY NUMBER OF SURVEYS - PANEL

BUYER BEHAVIOR BY NUMBER OF PANELS - PANEL



BUYER BEHAVIOR BY NUMBER OF SURVEYS – SOCIAL NETWORK

FIGURE 3, CONTINUED
BUYER BEHAVIOR BY NUMBER OF PANELS – SOCIAL NETWORK



Buying behavior of American sample sources compared by different measures of respondent activity (number of surveys completed, number of panels).

Now that we have a common language, the segmentations, we can have a common dialogue about the problems that haunt us.

BLENDING

We are dealing with dissonant sources and it takes a trained mind to comprehend how they go together. Social networks are vastly different from other online sources. Point system cultures, such as frequent fliers, are a stark contrast from those who are drawn from the social networks. Those within a demographic group who are online for the purposes of commerce are different from the many that virtually inhabit Facebook for social reasons. Their vantage points are different and their purchasing behaviors are in contrast.

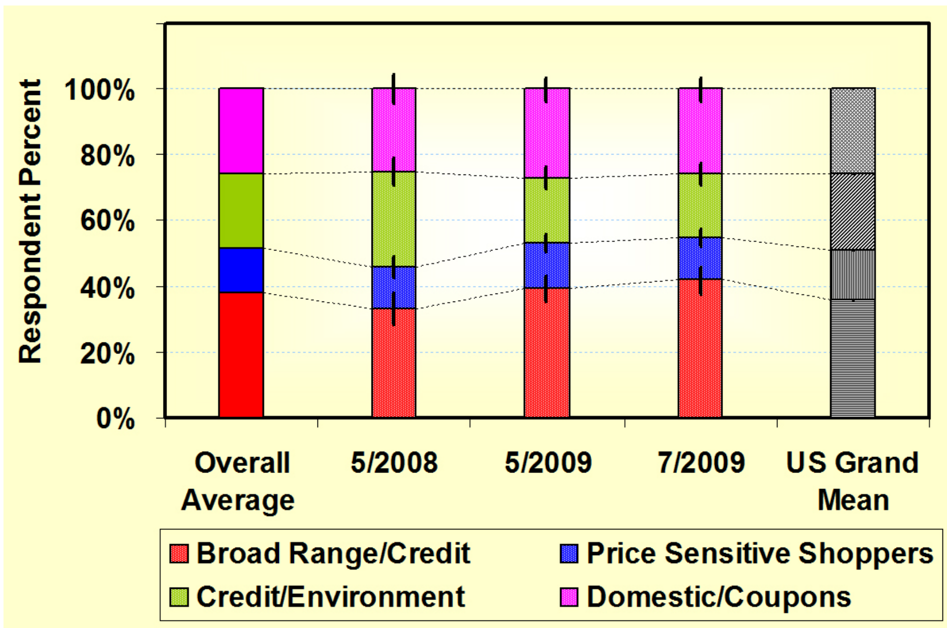
Blending of our sources is a critical epiphany for the industry. The music is disparate and we cannot ignore where it goes without addressing the issues. A haphazard mixing of sources is not a sample frame, it is just noise. Shortages of opt-in panelists that drive fear into our research pipeline cause us to move with neither scientific nor artful combination. We prefer to call this process "mixing." Blending, on the other hand, requires a detailed and transparent score that in itself is replicable, transparent, and by its definition, brings together the constituent elements in harmony.

RULES, RULES, RULES

We need rules – or we need the lack of them. Here we speak of sampling frame issues and beg for the combination of samples through techniques that create structure. We accept the bifurcation that the industry appears to be bent upon. There will always be consonant and dissonant sources. The mixing of social network, panel, and the scraping of information off the Internet walls will give us one source. We need a plan for both. This bifurcation will be unstable if the resultant data is incompatible. In the worst case, if the data leaves us swamped out by noise confounding its use to make decisions, it will be of trivial value and dangerous to our professional health.

The first hurdle in being able to compare our data is accounting for variability. We will be haunted if we cannot achieve a common language and an appropriate means to control it. Source related bias is a powerful driver but even within a given source, the segment distributions of the panels differ in a given time and can be dramatically different over time.

FIGURE 4
BUYER BEHAVIOR SEGMENT DISTRIBUTION PROFILE



Census balanced segmentations drawn from pro-profiled respondents during a one year period

Figure 4 gives the analysis of census balanced segmentations drawn from pre-profiled respondents drawn at intervals during a one year period. Most of the time, the segments track nicely. Sadly it only takes one divergence for the differences to be significant and meaningful.

It is not always necessary to match to a particular standard to evaluate the ability to replicate a sample. The historic segmentations of a sample source can be the standard that we go by. The variability of the segments can help diagnose heretofore unexplained changes in tracking data. Even ad-hoc studies benefit from the stability of knowing where the segmentations come from. Clearly, if the behavioral segments are unusually skewed, then the data collected and the decisions derived could be equally amiss.

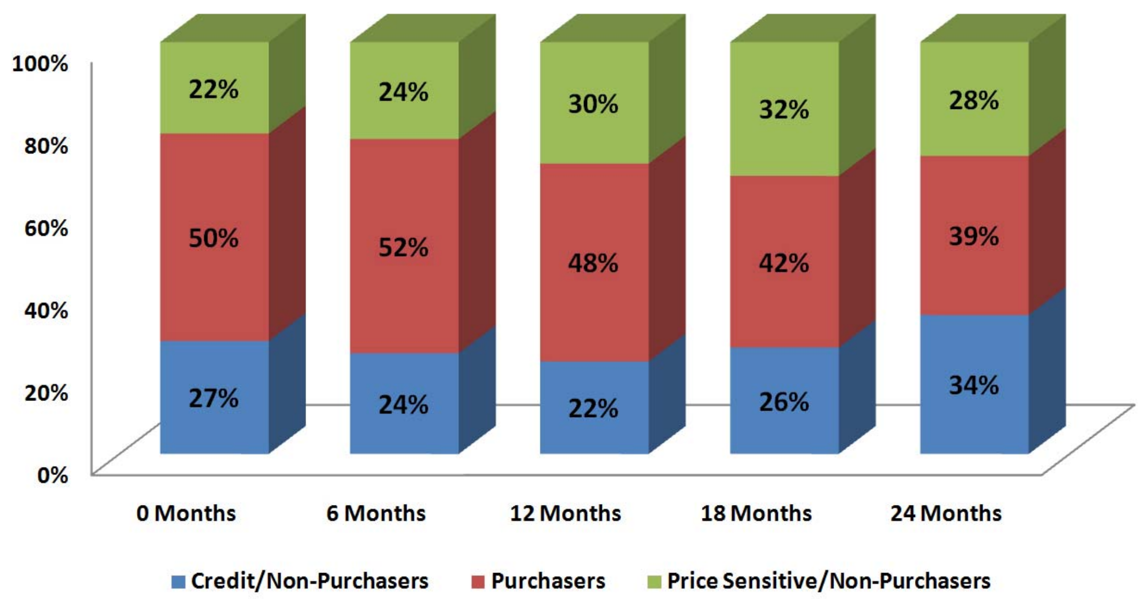
The AAPOR task force for online research invested considerable time scouring the landscape to assess what we know. It appears that with great reluctance, they concluded that there was a practical application for the method; "There are times when a non-probability online panel is an appropriate choice." (Baker et al, 2010)

Much of their reluctance to approve the method is founded in sources of bias. The absence of data on attrition bias, conditioning, and the differences between probabilistic and non-probabilistic sampling frames, feeds the uncertainty.

We can use our newfound common language and map the segment changes under certain conditions. For example, the potential for attrition bias is little understood. Joining a panel is parallel to joining any other activity. Many of us engage in activities in the hope that we will find what we seek in the new endeavor. However, when the actual experience is before us, we become disengaged and attrite. If we look at survey taking as a club where people exchange information, there are some who will warm to the activity more than others. Those who do not choose to continue may differ in some ways from those who do. In our language of segmentation where the segments are akin to behavior, the respondents who refuse to participate after a period of time, or a dose of experience may represent different behavioral groups and thus, by their absence, they will create a non-response bias.

Non-response bias is difficult to measure. On a practical basis, the ability to do so during a study is clearly problematic. A shift in the segment distributions represented by differing lengths of respondent tenure can give us a sense of how behavior changes as respondents attrite, or conversely, are retained by the interviewing system. (See figure 5.)

FIGURE 5
BUYER BEHAVIOR BY PANEL TENURE



Buyer behavior segment distribution changes with increasing panel tenure.

The possibility exists that there are other forces driving these changes. For one, Baker et al properly explains that conditioning effects are poorly documented in online panels. In addition, the changes we see may in part be driven by shifting demographics. We controlled for demography and found similar changes exist. Not all the variability is due to attrition, nor can it be fully explained by demography. The message is clear. We cannot explain the driving causes of variability. In the absence of an explanation, we must consider a solution.

RECENT INDUSTRY EFFORTS: REAL, LIVE AND ENGAGED

Our earliest efforts to improve online quality consisted of comparing online data to that of other modes. The subsequent chapter was characterized by efforts to verify the sources of variability. Academic concerns haunted commercial needs with a continuous hammering of data challenging the non-probabilistic character of the sampling frame. Most recently, the commercial sphere has directed its energies by cleansing the respondent population.

There has been quite a bit of effort expended on the differences between probabilistic and non-probabilistic sample frames. Important studies have emanated out of Stamford University that have focused on how the two sample types drive differences.

The tempest made around those sample frame issues is actually contrary to the overall tide. For the most part, research into online sources of error has been limited to comparisons with other modes, and the driving forces behind the variability within the data. Little data has been published that illuminates within panel consistency over time. Thus it is understandable that the most important advances towards better online quality have been respondent driven as opposed to sample frame driven. Here we speak to a cluster of initiatives that have gained favor in the most recent past. Peanut Labs pioneered digital fingerprinting in market research, a method now considered standard. MarketTools, with the support of Procter and Gamble and Microsoft, has made checking respondent data against outside databases "table stakes."

Relatively new to the scheme and championed by a number of others is the issue of respondent engagement. The thrust of these endeavors has resulted in a collection of indices that measure the degree to which respondents "satisfice" while completing surveys. The industry has also engaged in a mea culpa of sorts where structural differences (length, burdensome grids, repetitive questions, etc.) in questionnaires have begun to draw attention.

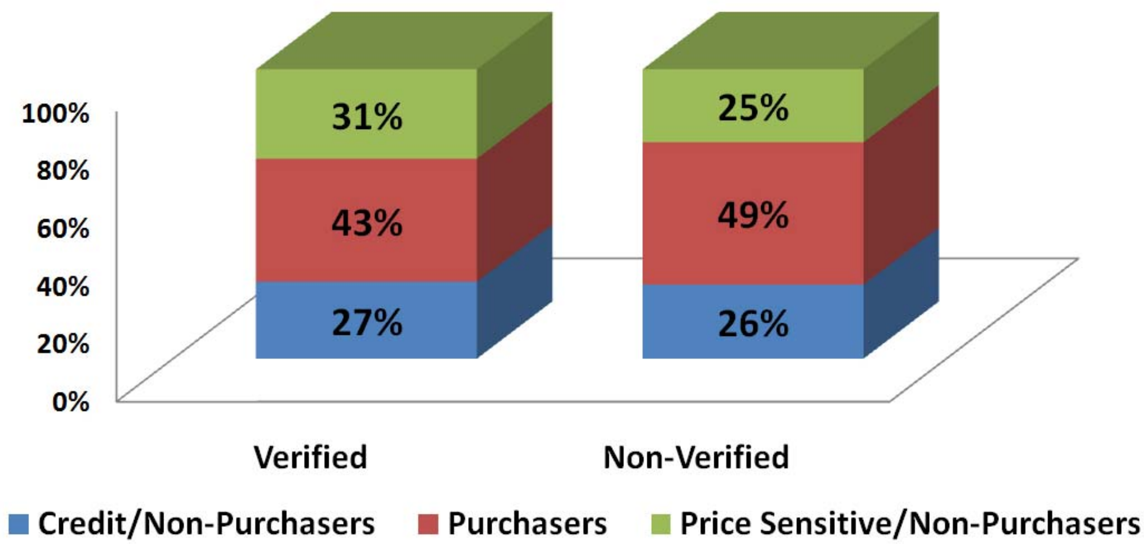
Considerable work, ours and others (including MarketTools) has been based on quantitative meta-analysis of executed surveys focusing on drivers of engagement. Disengagement is measured by any number of "bad" behaviors of respondents (speeding, straight-lining, satisficing, etc.). There have been some recent findings linking such behavior with other unacceptable and poor survey behaviors such as invalid and inconsistent responses (MarketTools, 2010). It is clearly possible to create a metric which captures the engagement level of the respondent without understanding why the engagement itself varies.

The current driving quality initiative in the United States appears to be: 1) the elimination of duplicate respondents; 2) verification that the respondent is real; and 3) measures of engagement.

For the most part, respondent verification identifies if there is such a person living at the address given by the respondent. A natural extension is to use outside data bases to verify the accuracy of other informational questions imbedded within the survey to test for respondent accuracy.

As we tag, or remove, those who fail to pass "real, live and engaged" tests, we will need to understand how the screen out of such respondents changes the nature of the sample frame and its ability to represent the targeted population. Figure 6 shows that differences exist between verified and unverified respondents.

FIGURE 6
BUYER BEHAVIOR BY VERIFICATION



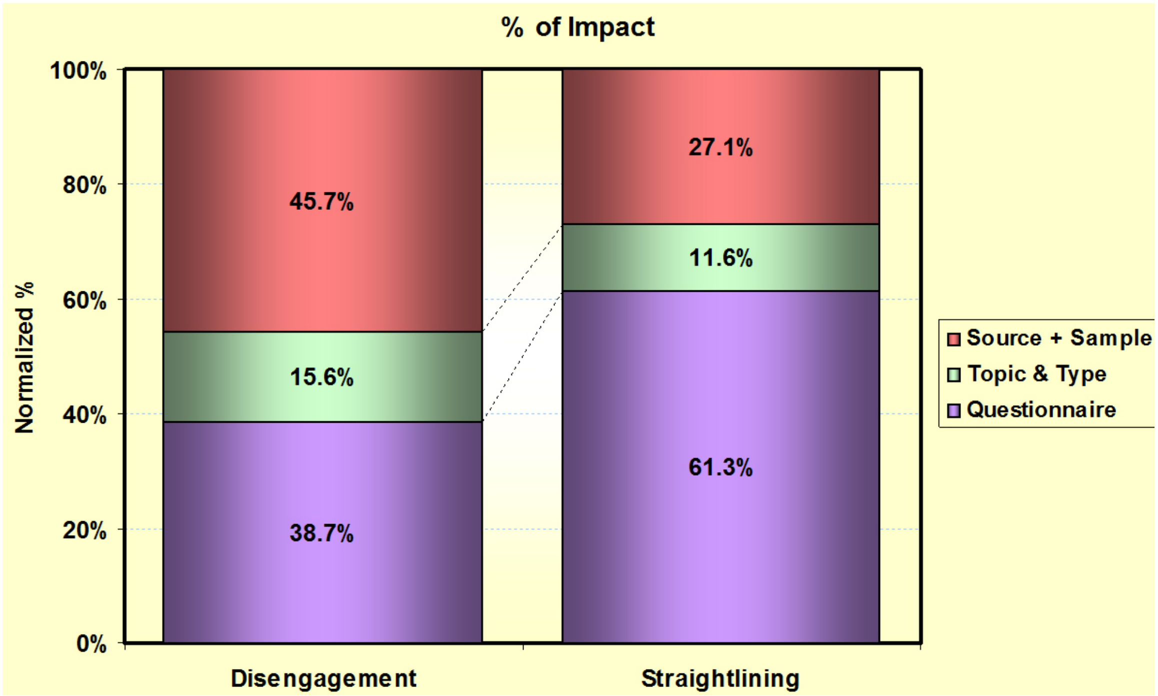
Provides data on differences between verified and unverified respondents.

For a while it appeared as if engagement would take the form of innovative methods of conveying questions to respondents. Various "flash" technologies along with creative animation all weighed in as the newest tool for creating a positive respondent experience.

Considerable effort has gone into measuring engagement levels. In addition, various tactics have been employed to better engage the respondent. Logically, increased incentives should influence engagement. In our research we examined 1,100 online studies, high levels of engagement existed in those highly incented groups. Our data confirms that of MarketTools (2010), which says that the drivers of engagement appear complex. In our analysis (figure 7), the affinity between subject matter and the source from which the respondent was derived both swamp out structural changes in the questionnaire.

There has been a maelstrom of debate over questionnaire design elements that influence respondent engagement. Again, shorter, entertaining questionnaires, without extensive grids should provide a better interviewing experience. MarketTools has constructed a predictive model that reports to score the engagement of respondents in advance of execution, allowing the researcher to make appropriate modifications.

FIGURE 7
CHARACTERISTICS OF SAMPLE AND SOURCE ARE ROUTINELY A DOMINANT FACTOR IN DETERMINING THE ENGAGEMENT OF RESPONDENTS



Various initiatives continue the concept of improving the respondent experience. The obvious benefit is that we need to retain the small population that is willing to feed us information. Their willingness improves response rates, once considered the critical measure in probabilistic sample frames, and should reduce attrition bias. These methods remain respondent driven as opposed to sample frame driven.

Research practitioners should govern the sample frame. What follows is a test project with the cooperation of Microsoft that shows how the sample frame can be matched and retained to a standard.

EXECUTION OF A BEHAVIORALLY PROFILED SAMPLE FRAME

In cooperation with Microsoft, a Utility Study was conducted using behaviorally profiled respondents. Six hundred census balanced interviews were completed between Friday April 8th and Sunday April 18th, 2010. A soft launch of 49 interviews was completed on Friday and Saturday April 9th and 10th. Data collection was paused to evaluate the first 471 interviews that were completed on April 15th, 2010. The average length was approximately 15 minutes.

Each respondent was pre-profiled using a standard profiling questionnaire. The questions were selected to allow a consistent standard and independent assessment of the respondents in a sampling frame or a completed study set. Two data bases have been created – one from the Grand Mean Project® (about 100,000 respondents in 35 countries), which generated samples from 20 American online sources, mainly double opt-in panels. The second is a pool of respondents provided by a number of online panels that were behaviorally profiled through the standard questionnaire. These respondents are now part of a growing pool.

For this assessment, at the end of the third quartile, 2,813 potential respondents were invited to complete the Utility Study. To compare these respondents to the Grand Mean (GM), we took a census balanced subset of 2,000 GM respondents, who had also completed the profile questionnaire and used them as a standard.

A second analysis was completed on Monday April 19, 2010 after 600 interviews had been completed. It is standard to analyze results after a soft launch, and then at every quartile of a study to completion. We shall report here on the results of the third quartile and completion analysis.

Quotas for age, income, gender and race were employed to demographically balance the final population of completed interviews.

FINDINGS

The Buying Behavior segmentation consists of 37 input variables broken into three segments. Two of the segments are non-purchasers of most of the goods and services tested and the central segment is the predominant purchasing segment. The titles of the segments in all our segmentations reflect the strongest loading variables making up the segment.

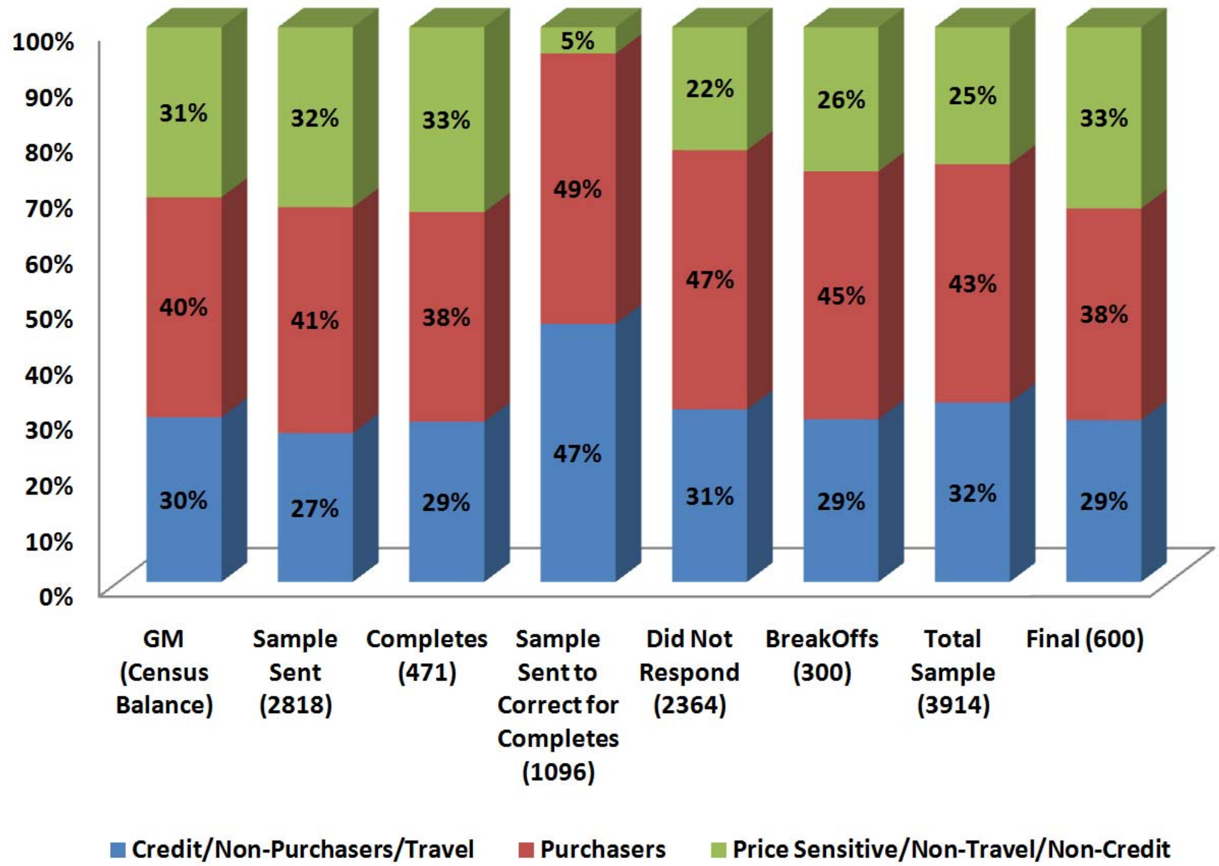
Figure 8 illustrates the comparison between the buyer segment distributions for the Grand Mean - Census Balanced and a sequence of analysis drawn as the Utility Study progressed.

The distributions in figure 8 begin with the Grand Mean standard (first stack), followed by the 2,818 respondents recruited (second stack) to complete the first 471 completed interviews (third stack). These three are not significantly (χ^2 P>0.05) different.[1)]

The fourth stack represents sample respondents that were e-mailed as an augment sample to complete the study. At this point, when looking at the demographic balance, the younger, lower income and non Caucasian respondents were lagging behind and a disproportionate sample send was required to correct for the demographic imbalance. The fifth stack includes the non-responders in the study – those who simply ignored the e-mail requests. This grouping has a very large central purchasing segment at 47%, very different from the targeted 40% in the Grand Mean standard. The sixth stack includes those respondents who broke off after beginning the questionnaire. This stack (the break-offs), while statistically different from the non-responders, is most similar to those that failed to respond than those that completed the survey. The seventh stack includes the total sample sent, with segment distributions extremely different from the final distribution of the completed interviews.

The shift among the segments is towards an increase in price sensitive purchasers, a decrease in the principal purchasing segment and a decrease in the credit, non-purchaser segment.

FIGURE 8
PROGRESSION OF BUYING BEHAVIOR SEGMENTS DURING THE PROGRESSION OF A UTILITY STUDY



DISCUSSION

The pre-profiling proved to be a useful tool for balancing behaviors in "real-time". Respondents solicited and those who failed to respond differed from those who completed the Utility Study. The study group selected was de-duplicated, outside verified, tested for engagement and the accuracy of their responses. They would be considered to be a "preferred" online population. The difference between responders and non-responders is considered to be a response bias. In this instance, the sample pool of some 50,000 respondents had been previously profiled through a standard questionnaire and then solicited for a second, "Utility" research study. Those who responded to the Utility Study tended to be older, better educated, and from economically higher strata. The non-responders came from buying, sociographic and media segments that involved more Internet usage and were more likely to purchase a variety of goods. The older responders were more often frequent survey takers with a higher predilection to belong on multiple panels. The non-responders exhibited shorter tenure on online panels. The non-responders are by some measures, less engaged in the survey taking process.

Demographic quotas provided scant control over variability in the array of behaviors measured through a set of multivariate segmentations. The non-responders appear to represent those most likely to abandon the online panel culture and exhibit this tendency by their failure to respond to this study as well as a suite of attributes including their demography, behavior, segmentation alignment, and appetite for questionnaire consumption. Those that broke off mid-questionnaire are more similar to those that failed to respond than they are to those who completed the research. We hypothesize that the non-responders are likely similar to people who leave the panel system.

The net result is that the distribution of segments among the sample drawn is different from the completed interviews. This difference is the product of midterm break offs, those who failed to respond and initial break offs. All sources of non-response had similar effects. Primarily, the purchasing segment was more heavily represented in the non-responders than in the final sample and the price heavy purchasers less represented. The mid-term correction sample send was made possible by segmentations collected in advance.

In Market Research, when a tree falls in the woods and no one is listening, it does make a sound. By pre-profiling all respondents, we were afforded the ability to balance behavioral differences that arose even when demographics were in balance. This changes how we might practice online research in the future. We can expect that those who respond to our questionnaires are different from those who fail to attempt the screener. In turn, the screening process eliminates respondents that might otherwise have represented a different and valued set of opinions. We can no longer assume that quota controls on the usual battery of demographic variables will suffice. We should routinely measure non-response bias. The value of doing so is the difference between an unknown sample frame and one that we can understand and adjust.

Finally, the respondents who answered this questionnaire are sending us a message. It appears that the non-responders are those who quit in the screener or the questionnaire. In fact, the panel members who simply opted-out from any level of participation are the best candidates to abandon panels in general. This is a poignant clue as to the nature of change in online panels in general.

CONCLUSION

The sample sources provide us with the notes from which we draw our music but it is our role as conductor, composer and listener that will bring us to the next horizon: focusing on the sample frame instead of the sample alone. We are the keepers of that frame. It is our expertise that makes so many voices in a crowd a representative opinion. If there was ever a true probabilistic sample then god bless its demise. We are the new jazz.

Jazz in the United States took its origins from other cultures. Its freely ranging style clearly represents a state of emancipation through dissonance. The concept of a freely spirited mind voiced through its music is global: as such so is our work. We are a global business with a responsibility for global ideas. The end users of our research are global companies who seek minds that can make good music. Thus the solution of the sample frame must be global in its application. Getting and maintaining meaningful (representative), quality (proper responses) and consistent online samples must be based on global standardization of sources.

The idea of dissonance has always been there. The strain between two ideologies that differ causes discomfort. As time goes on we become accustomed to the dissonant element, find value in its contribution and, if we are fortunate, driven to incorporate it into a framework of beauty. Jazz too has changed over the years but the influence it has had on music will endure.

We propose the use of profiled respondents and structural segmentations as a common language. The orchestration of standards to achieve these aims will require considerable standardized information about individual respondents. But further, we must understand the models that we choose to employ. If there is a lesson that we must take from the past it is that our industry relied on models that enjoyed our confidence. Nothing has changed; we just have to make new music.

FOOTNOTE

1. The measure of differences between the distributions is based on the maximum of the differences of each of the distribution elements divided by its standard error. BBC. (2006, 7 March). The MySpace Age [Electronic Version] from http://news.bbc.co.uk/1/hi/magazine/4782118.stm

REFERENCES

Baker et al., March 2010. AAPOR Report on Online Panels. Prepared for the AAPOR Executive Council by a Task Force operating under the auspices of the AAPOR Standards Committee, p.52.

MarketTools white paper. 2010. When Good Respondents Go Bad: How Unengaging Surveys Lower Data Quality.

Walker, Robert, Raymond Pettit, and Joel Rubinson. 2009. A Special Report From The Advertising Research Foundation: The Foundations of Quality Initiative, A Five-Part Immersion into the Quality of Online Research." Journal of Advertising Research 49: 464-485.

THE AUTHORS

Steve H. Gittelman is President, Mktg, Inc., United States.

Elaine Trimarchi is Executive Vice President, Mktg, Inc., United States.