

# White Paper



---

## Survey Lockouts: Are we too cautious?

*By Philip Garland, Ph.D., SSI, Debra Santus, Ipsos & Reena Uppal, Ipsos*

There is a general belief in the commercial survey research industry that responses are affected when people participate in multiple surveys in close succession, especially when the topics of the surveys are similar. For this reason, “lockout” periods, or category exclusions ranging from one to six months or longer, are often practiced by various research firms to prevent what is commonly called “conditioning.” The primary factor that is thought to induce conditioning is the amount of time lapsed between two survey events similar in content. Conventional wisdom holds that the second surveyor would like a “clean slate,” whereby the responses to his questionnaire are free from contamination by prior survey participation. As the industry addresses its biggest challenge—the shortage of respondents needed to take surveys—a re-examination of conditioning effects is timely.

Some research has shown that participation alone, regardless of the survey topic, can affect behavior intention. For example, Teal and Haechrel (2008) showed that people who were frequent survey takers reported lower propensities to purchase new financial products—even when the survey participation occurred in other categories. Conditioning, then, is thought to be a kind of frequency effect whereby participation regularly in a category of products affects subsequent survey responses. For example, reports of brand awareness may depend on whether that brand was mentioned in a recent survey undertaken by the respondent.

Though there is neither consensus definition of conditioning, nor the factors that contribute to its onset, some research has shown that participation frequency within a category can yield differences in certain kinds of survey responses. Grover and Virens (2006) showed that ratings of products’ uniqueness decrease over time when participants are asked to rate several new product concepts in close succession. Chandon, Morwitz, and Reinartz (2005) showed significantly higher correlations between reported intention to buy a product and actual purchasing behavior among people who were surveyed compared to people who were not questioned. Cartwright and Nancarrow (2006) showed that responses collected after recent and frequent completion of a questionnaire become more dissimilar. Segers and Franses (2007) demonstrated that people’s ability to recognize newspaper headlines declines when they are asked to do so frequently and in close succession.

The findings of these studies focused on frequent survey participation in a given category beg a question about validity. If responses yielded by one survey depend on the recency of participation in previous surveys, then which of two (or more) responses to any one question represent the actual attitude or behavior of interest? Little if any research has attempted to determine whether distributions of responses that are more similar or dissimilar after conditioning is the more valid, and thus desirable, outcome.

Whatever the optimal outcome of responses yielded by people who have completed an identical questionnaire several times, the analytical strategy of researchers interested in conditioning has been, to this point, limited. Research to date concerned with conditioning for the purposes of category exclusions or lockouts, has focused almost exclusively on the changes in responses to one questionnaire over *n* waves of administration. By doing so, researchers interested in understanding conditioning have operationalized the laboratory setting in a way that fails to replicate the conditions of the industry procedure for lockouts. Lockouts have been used to prevent a person from taking a survey in the identical category as a recent survey previously taken. Of course, the questionnaires of any two researchers are rarely, if ever, identical and researchers desirous of lockouts are concerned

*(continued)*

about identical categories, not identical questions.

Therefore, rather than test the repeatability of responses in identical questionnaires, explorations of conditioning should be undertaken using similar questions within a particular category. In this light, this paper departs from previously employed designs and attempts to more faithfully replicate the setting in which category exclusion is thought to prevent conditioning.

With this design in hand, the analytical strategy concerning conditioning effects ought to examine whether the responses to the most recent questionnaire change across groups who have taken surveys in close succession or occasionally. If conditioning indeed has undesirable implications for data quality, then the responses to similar questions in a given category should be more different from one another and less able to predict reasonably related behaviors among respondents who had recently completed a survey in the category than those who had not.

Our exploration, then, is concerned with illuminating a) the effects of conditioning on data quality, and b) the minimum time lapse required to induce such effects. A second factor that may interact with conditioning effects is the types of questions asked. In other words, conditioning may occur for some categories but not others, so researchers can decide whether to be more specific in the categories deserving of lockouts—if any. It appears worthwhile to study conditioning within three broad question domains—product ratings, behavior recollection, and socio-political attitudes.

## Method

The study used a 3X3 (incidence frequency X topic), mixed design whereby respondents completed similar questionnaires within categories at varying intervals—monthly, bi-weekly, and weekly. The high frequency group participated in weeks three, four, and five. The medium frequency group participated in weeks one, three, and five. The low frequency group was surveyed in weeks one and five. A control group was surveyed only in week five.

We measure conditioning effects in three ways:

- 1) correlations between items,
- 2) predictive validity, and
- 3) comparisons of responses.

## Sample

To examine conditioning effects, we arranged for three questionnaires to be administered to two large national samples of U.S. adults by the survey research firms SSI and Ipsos. Panelists were initially recruited by the firms through RDD, invitations via web newsletters, and web advertising based recruitment where panelists opted to participate in surveys. All of the panelists provided their e-mail addresses to receive invitations to participate in surveys run by the firms. Upon agreeing to become members of the panels, respondents answered a variety of demographic questions.

*(continued)*



Interviews were conducted via the Internet. The samples were nationally representative created by a technique in which panelists are matched to characteristics (gender, race, education, age, and income) of the U.S. census. Respondents were randomly assigned to receive surveys at varying intervals. Data collection began on November 11, 2008 and was completed on December 13, 2008. The invitation, delivered by email, described the content and importance of the study, and included a unique hyperlink authenticating the respondent's participation in the study. A total of 7,070 people participated in the study across the samples produced by both firms.

## Measures

*When a new technology or product is invented that interests you, how soon after its release do you purchase it?*

Immediately, Very Soon, Somewhat Soon, Not very Soon, Not At All Soon.

*How important to you are brand names when purchasing an item?*

Extremely Important, Very Important, Somewhat Important, Slightly Important, Not At All Important.

*How often do you read a newspaper either in print or online?*

Always, Very Often, Sometimes, Rarely, Never.

*Do you have cable or satellite television?*

Yes, No.

*During the past year, how many times have you recommended a cable news channel to someone?*

*Do you have a mobile/cell phone?*

Yes, No.

*How often do you use a cell phone?*

Always, Very Often, Sometimes, Rarely, Never.

*During the past year, how many times have you recommended a cell phone service provider to someone?*

*In your opinion, which company offers the "coolest" portable music player (e.g., an Mp3 player)?*

Apple iPod, Microsoft Zune, Sony Walkman, Philips, Samsung, Other, Don't Know.

*Do you have a portable music player (e.g., an Mp3 player)?*

Yes, No.

*During the past year, how many times have you recommended your brand of portable music player to someone?*

*During the past month, how many times have you recommended a clothing store to someone?*

(continued)



*Have you heard of the Gillette Power Fusion razor?*

Yes, No.

Respondents who said yes were asked: “Have you tried the Gillette Power Fusion razor?”

Yes, No.

*Respondents who said no to either question were then asked: “Are you willing to try the Gillette Power Fusion razor?”*

Yes, No.

*Please list below the brands of soda you’ve heard of.*

*In your opinion, which brand of soda is sold the most in the United States?*

*Have you heard of Jones Soda?*

Yes, No.

Respondents who said yes were then asked: “Have you tried Jones Soda?”

Yes, No.

Respondents who said no to either question were then asked: “Are you willing to try Jones Soda?”

Yes, No.

*Do you support or oppose affirmative action policies that give preference to African American applicants applying for jobs?*

Strongly Support, Somewhat Support, Slightly Support, Neither Support nor Oppose, Slightly Oppose, Somewhat Oppose, Extremely Oppose.

*Do you support or oppose affirmative action policies that give preference to female applicants applying for jobs?*

Strongly Support, Somewhat Support, Slightly Support, Neither Support nor Oppose, Slightly Oppose, Somewhat Oppose, Extremely Oppose.

*Have you voted on an Affirmative Action law in your state?*

Yes, No.

## Results

To replicate the most basic findings of previous research investigating conditioning effects, we compared the distributions of responses to two questions designed to capture brand recognition between the treatment groups. This study replicated the findings of previous research: people who take surveys that contain brand names are more likely to report familiarity in subsequent survey administrations containing the brands. Moreover, the effect was more pronounced for the question containing the “obscure” brand—the one with substantially smaller market share compared to its competitors.

*(continued)*



These results also confirmed that respondents may learn how to take surveys more efficiently when they are exposed to more of them. Specifically, respondents who took more than one survey were more likely to offer 'Don't Know' responses to several questions.

Despite these two findings that seem to support a policy for locking out panelists, these data also suggest that there are few differences between people who take many surveys in close succession and those who do not. The distributions of responses across 14 of the 18 questions were statistically identical among all groups, including the control group. In other words, whether the questionnaire was the only one completed or the last of a series of surveys containing the same topics, the responses were the same for nearly all of the questions in our study. While distributions of responses shed some light on comparability of responses, we explored whether correlations between items that are theoretically assumed to be related were in fact linked.

Using OLS regression, we tested whether five items concerning brand awareness, electronics ownership, and brand importance predicted responses to the question: "When a new technology or product is invented that interests you, how soon after its release do you purchase it? Immediately, Very Soon, Somewhat Soon, Not Very Soon, Not At All Soon.

The R square values increased incrementally among high participation frequency groups (Control, .113; Monthly, .241; Bi-weekly, .252; Weekly, .322). That is, as people took more surveys, the ability to predict a face valid relationship increased. This finding suggests that as people become more familiar and comfortable with a category under research, their answers to related questions become more sensible—a result counter to the prevailing notion that effects of conditioning are exclusively harmful to data quality.

## Discussion

Changes in data as a result of conditioning are difficult to find. There were few differences in the answers given as a result of repeated exposure to the same survey. It is important to note that this study exaggerated the situation which would normally be found in a panel environment, by exposing respondents to the same survey within a very short time span. Even with this exaggerated exposure, the impact was minimal. Since the impact of exclusions on the universe of available respondents is such a potentially serious threat to the industry due to the lack of available respondents, lockouts should only be undertaken if there is concrete evidence of the benefits received to the quality and validity of the data for the specific research project being undertaken.

Despite the evidence demonstrating effects of survey participation on brand recall for obscure brands, some research suggests that people are unable to recall the topic of the surveys they took in the past (Cape, 2008). Just 35 percent of respondents reported having the ability to recall the subject matter of a survey taken one to two weeks prior while 21 percent of respondents reported being able to do so one month after participation. To be sure, people may over-report their ability to recall subject matter in order to appear engaged or even claim superior cognitive skills. However, if pressed to report the actual subject or in a comparison to a factual record of behavior, the percentage of people who actually remembered the topic of a prior survey is likely lower than that yielded by the self report. Findings such as this, in conjunction with the data reported in this paper, suggest that conditioning is

*(continued)*



at least confined to specific instances.

It is important to note that sample sources for this study were two large online access panels, and therefore respondents are likely to have taken other surveys within this environment, leading to some underlying “environmental” conditioning effects in all the groups. Moreover, it is widely accepted that any given panelist may be a member of more than one panel, thus this study was unable to account for the effects of participation within the category outside of the bound of the study itself.

Future research should investigate how conditioning interacts with panel tenure. Because panels rely heavily on available members, there may be a baseline level of conditioning that all panelists exhibit once they have been a member of a panel for a period of time—in other words, a floor effect. Other research might investigate whether conditioning occurs in other substantive categories. Lastly, future work ought to explore conditioning in new product testing due to the inherent sensitivity of concept testing.

*(continued)*



## Appendix of Detailed Results

### **Gillette Power Fusion brand recognition:**

Mean: 0.82 (had not heard of product = 0; had heard of product = 1)

80% had previously heard of the Gillette Power Fusion.

$x^2 = 12.04$

p-value = 0.007

The p-value indicates a deviation in the data due to an outside influence. This presents the case that there is an effect on succeeding surveys of panelists in the high frequency sample group.

### **Jones Soda brand recognition:**

Mean: 0.44

45% of the sample had heard of this product.

$x^2 = 4.011$

p-value = 0.26

There is no significant difference between expected and observed results. There are no stimuli skewing the data as was the case with Gillette Power Fusion.

### **General opinion of products:**

“Don’t know” was the first or second most selected answer, except for the question on highest quality cable news channel.

$x^2 = 34.31$

p-value = 0.003

### **“Coolest” music player:**

“Don’t know” was the most frequent answer (almost 50%)

iPod was the second highest answer (38.4%)

$x^2 = 35.65$

p-value = 0.008

### **Stores with best reputation:**

Macy’s = 21%

Don’t know = 20%

$x^2 = 48.08$

p-value = 0.002

All three of these questions had  $x^2$  and p-values indicating significant differences between the gathered data and the expected data, implying possible conditioning effects due to taking multiple surveys in close succession.

*(continued)*



This however is not the case with the final question in this series on the best shaving razor. Fully 56.8% of respondents answered Gillette and 20% (the second most popular answer) answered don't know.

$$x^2 = 16.61$$

$$p \text{ value} = 0.165$$

There is no difference between expected data and what was collected.

**Use of cable or satellite television:**

Mean = 0.91 (0 = no, 1= yes), (91% of respondents do own cable or satellite TV)

$$x^2 = 8.99$$

$$p\text{-value} = 0.029$$

**Owners of portable music players:**

Mean = 0.96

$$x^2 = 23.336$$

$$p\text{-value} = 0.001$$

Both p-values show that there is a significant difference between expected and actual data.

**Own a mobile or cell phone:**

Mean = 0.88 or 90%

$$x^2 = 2.72$$

$$p\text{-value} = 0.437$$

There is no statistical significance.

**Voting on Affirmative Action:**

Mean = 0.21 (21% of panelists have voted on Affirmative Action in their state)

$$x^2 = 0.296$$

$$p\text{-value} = 0.961$$

There is no statistical difference from what was anticipated.

**Tried Jones soda:**

Mean = 0.48 (0 = no; 1 = yes): nearly 50% of panelists have tried the beverage.

$$x^2 = 5.95$$

$$p\text{-value} = 0.026$$

The data is being affected by the variable.

**Have tried Gillette Power Fusion:**

Mean = 0.26

$$x^2 = 3.6$$

*(continued)*



p-value = 0.73

The two subsequent questions asking whether or not the panelist would be willing to try each above product have high means of 0.82 for those willing to try the shaving razor and 0.81 for those willing to try the beverage. Their chi-squared values and p-values are 4.22 and 0.65 for the razor and 5.95 and 0.429 for the beverage. All three indicate no statistical significant difference in the data.

Data on questions relating to daily lives show little difference between assumed and gathered data. 38% of participants responded that they always read the newspaper either in paper or

online form despite the mean, 2.8, which corresponds to “very often” and “sometimes,” with inclination towards the latter.

$\chi^2 = 20.18$

p-value = 0.064

There is no statistical difference.

When asked about cell phone usage, 35.9% of respondents answered that they always used their cell phones, while the mean was 2.99 (2 = very often, 3 = sometimes).

$\chi^2 = 30.05$

p-value = .012

36% of panelists responded that they purchase new products somewhat soon and 35% responded not very soon.

Mean = 1.67 denoting that very soon and immediately were the most chosen answers.

$\chi^2 = 16.28$

p-value = .179

The final question in this series asked the importance of product brand names.

49% of respondents said that brand names are somewhat important, with a mean of 2.1.  $\chi^2 = 13.88$

p-value = 0.308

There is no statistical difference between the found and collected data sets.

Statistics for all three socio-political questions illustrate that we can accept the data as unaltered by outside effect.

The first asked respondents their opinion of Affirmative Action policies that give preference to African Americans when applying for a job.

*(continued)*



31% replied 'neither support nor oppose' with the mean of all responses at 3.54 (3 = neither support nor oppose; 4 = slightly oppose).

$\chi^2 = 12.00$

p-value = 0.85

The second question concerned opinions of Affirmative Action policies that give preference to females when applying for a job. 32% said 'neither support nor oppose.' The mean was 3.30, (3 = neither support nor oppose; 4 = slightly oppose).

$\chi^2 = 14.196$

p-value = 0.716

33.8% of panelists designate their views as moderate. The mean, 2.44, supports this percentage (2 = somewhat liberal; 3 = moderate).

$\chi^2 = 17.68$

p-value = 0.279

The p-value indicates no interference with the data by previous exposure.