

EXHIBIT 2

**UNITED STATES DISTRICT COURT FOR THE
EASTERN DISTRICT OF NEW YORK**

**BARBARA SCHWAB, et al., Individually
and on Behalf of All Others Similarly
Situated,**

Plaintiffs,

v.

PHILIP MORRIS USA, INC., et al.,

Defendants.

Civil Action No. 04-1945 (JBW)

**SUMMARY OF EXPERT OPINIONS OF
DR. WAYNE S. DESARBO CONCERNING
THE JULY 14, 2006 RESPONSE OF DR. JOHN HAUSER**

**SUMMARY OF THE EXPERT OPINIONS
OF DR. WAYNE S. DESARBO**

I have been hired as an expert to comment on the work performed by Dr. John Hauser on behalf of the plaintiff in this court case. I assume that the data provided to me from the two Hauser studies is accurate and correct in representing the data collected from both studies. My work is ongoing and I may update and/or revise his results and conclusions as I review additional data and information. Based upon my review of the materials and my expertise in Marketing, Statistics, and Psychometrics, I have formed the following opinions, all of which are stated to a reasonable degree of scientific certainty. In this document, I will concentrate on the recent Hauser Response submitted to the court on 7/14/06, and only to those responses in that document that pertain to my original criticisms contained in the document I submitted to the court back in May 2006. I refer the reader as to my credentials and my summary of the Hauser study contained in my initial May 2006 critique that I incorporate by reference here.

**I. SUMMARY OF MY ORIGINAL CRITIQUE OF THE
HAUSER CONJOINT STUDY**

In my original criticism submitted to the court back in May 2006, I denoted the following 26 problems associated with this initial Hauser research. The asterisks that are placed next to the majority of the 26 comments I list below designate those areas of my previous criticism where no explicit response was made in the July Hauser document.

1*. The Greenfield Online Database/Panel utilized by Dr. Hauser is restricted only to consumers that have Internet access and volunteer to participate in such studies.

According to the Greenfield web site, Internet access is estimated at approximately 75%

of the population. Thus, 25% of the population is not reachable with use of this panel. Such statistics regarding the accessibility of the Internet to smokers of light cigarette brands is not available. The pertinent issue here is whether this 25% of the non-represented portion of smokers of light cigarettes have different utility functions, preferences, and choice patterns than the rest of this population (not just differing demographics). Thus, all estimates derived from such restricted sampling may be biased, and Dr. Hauser has not demonstrated that this potential bias is insignificant.

2*. In addition, there may be additional bias introduced as a function of respondents' willingness to participate in this study. Of the 52,402 invitations sent out to panel members, 44,159 refused to participate. That is, 84.3% of this sample refused to participate in this survey and that fact may lead to selection bias despite the efforts made for re-contact; only 15.7% agreed to participate. Dr. Hauser fails to demonstrate whether this 15.7% is representative of the target population of light cigarette brand smokers with respect to their underlying part-worth utility functions, preferences, and choices (not just demographics). Are there such differences with respect to those panel members that refuse to participate?

3*. Consumer panels are often utilized in Marketing as a convenient manner to collect information from consumers. Care, however, must be given to ensure that such panels are representative of the behaviors, characteristics, attitudes, preferences, choices, etc. of the target population of interest. Given that the Greenfield panel consists of professional survey takers, it is unknown how representative they are of the total population of light cigarette brand smokers. Dr. Hauser has not demonstrated that the light cigarette brand smokers participating in this Greenfield panel are truly reflective of the various smoking

behaviors, attitudes, preferences, perceptions, feature importance salience, etc. (not just demographics) of the true population of light cigarette brand smokers. No evidence of this important aspect has been provided in the initial Hauser report.

4. As mentioned in the summary statement in the first section of this report, the Hauser study utilized a quota sampling procedure in an attempt to match the national demographics of the *entire population* with respect to age, gender, income, and geographical region. However, even assuming demographics are the appropriate basis for a stratification of the entire national population, the matching process should have been done with respect to the national demographics *of light cigarette brand smokers*, which Dr. Hauser failed to do (see footnote 12 on p. 13 of his initial report which admits to this potential distortion). My Exhibit II in my initial report displayed the discrepancies between the sample characteristics (marginal distributions of the four demographic quota variables) of the respondents of this study in comparison with the U.S. population characteristics (U.S. Census Bureau) and that of light smokers as taken by the 2002 NCS (National Consumer Survey) study. Chi-Square tests were performed for each of these four demographic variables to examine whether the Hauser distributions were statistically the same as the US population (from the US Census data) and the NCS national study for light cigarette brand smokers. As shown in that Exhibit II, with the sole exception of gender in the NCS study, the Hauser sample deviated significantly from **both** distributions for **every other** demographic quota variable ($p \leq 0.01$). In addition, there are significant differences between the distributions for these four demographic variables between the US population (census data) and the NCS light cigarette smoker data indicating the fact that light cigarette brand smokers **do differ** demographically from the

overall US population characteristics. Given the discrepancies here between the three distributions, it is not clear which population the Hauser study can be properly generalized to. As such, Dr. Hauser has not demonstrated that his conjoint analysis survey sample is representative of the class of light cigarette brand smokers or that his results can be scientifically generalized to this same class.

5*. The use of these demographic variables as a basis of setting sample quotas is very questionable in this particular application. Ideally, one selects consumer characteristics that are relevant to the aims and objectives of the study. Without prior extensive theory or empirical evidence, one has to demonstrate the connection here between demographics and the preferences, behaviors, utility functions, and choice decisions that light cigarette brand smokers make, which Dr. Hauser has failed to do. The use of demographic information is convenient to Marketers for constructing samples, but may not be very much related to the preferences, behaviors, utility functions, and choice decisions that light cigarette brand smokers make. I explored the impact of these four demographic characteristics on the derived conjoint importances that were computed for pack, taste, health risks, and price to investigate the strength of these relationships. As shown in Exhibit III's MANOVA analysis from my initial report, these demographics, while somewhat significant in saturated model form for three of the four importance measures, have very little impact in fully explaining the derived importance of these four conjoint factors in light cigarette brand smokers, accounting for between only 11-17% of the total variation in these importances across the population, with the saturated model including all main effects and interaction terms providing adjusted R-squares range from 0 - 0.064 which are all trivial in magnitude. None of the main effects are uniformly significant at p

≤ 0.05 . Thus, the four demographic characteristics utilized by Dr. Hauser as a basis for the selection of his sample have little relevance to the purpose of this study, and it is therefore very questionable to use them as a basis for setting quotas for sampling.

6*. In addition, although the potential respondents were selected at random from the Greenfield Online database to match the 72 demographic cells of the national average, the respondents were screened via quota sampling so that once any particular cell was filled, no further respondents who matched that description would be allowed. While efficient, the net result is **not** a probability sample per se, and thus one can not calculate confidence or precision statements about the sample estimates that would generalize to the entire national population, and much less to the national population of light cigarette smokers (cf. Green, 1988). In other words, there is no way to calibrate how accurate the sample projections are to the population being studied. I therefore question the *external validity* and generalizability of this study based on all these issues discussed above.

7. There was other important demographic information (cf., Carter and Silverman, 2004) that was not utilized by Dr. Hauser in setting quota levels for the sample in this particular study. For example, ethnicity, occupation, and employment status were **not** utilized in the sampling frame of this study and are demographic characteristics that may have higher significant relationships with aspects relating to light cigarette brand purchase and consumption than do many of the demographic characteristics utilized in Dr. Hauser's study. In fact, an auxiliary data file suggests that Hauser evidently measured ethnicity, occupation, and employment status for a subset of these 627 respondents. Given the vast amount of missing responses per variable, a fully saturated MANOVA was not possible to estimate. However, in looking at the relationship between

these three variables and the four derived importance measures investigated previously via MANOVA, we see higher R-square fits than we witnessed in our previous Exhibit III for three of the four importance measures. Exhibit IV in my previous document presented the MANOVA results estimated with these three factors. Two of the four models are significant, and the corresponding R-square (ranges from 0.440 – 0.631) and adjusted R-square values (ranges from -0.079 – 0.440) are higher on average than the four demographics utilized for the sampling quotas. Thus, it can be legitimately argued that the sampling selection criteria utilized was incomplete and should have included additional demographic characteristics that together could explain more of behavioral phenomena or derived estimates. Even so, demographic information alone does not satisfactorily explain variation in the estimated conjoint importances.

8. Given that such quotas were not set on these other factors, Exhibit V in my previous document demonstrated that the resulting sample does not generalize to either the complete US population or the light cigarette brand smoking population with respect to two of these three other demographic factors for which we could align response categories (we could not match up the occupational categories for the three studies, given the drastically different response scales utilized). Thus, the Hauser sample was strongly biased and is not representative of either the US population or the NCS light cigarette smoker sample.

9*. The decision to smoke cigarettes is a complicated process and involves a series of decisions whose complexity and sequence may differ depending upon the individual. Smokers confront several decisions, including the type of cigarettes, attributes, brand, and consumption level. Such nested decisions most likely depend upon different factors

such as family history/acceptance, peer pressure, health concerns, taste, price, life stress, the relevance of the brand image of those brands in the smoker's consideration set, etc., and the sequence might vary upon the individual. One major problem associated with the Hauser study is that it had absolutely nothing to do with these specific real-world decisions. Nowhere in the study is there any explanation for these types of decisions. Indeed, it appears that the primary objective of the study listed on page 4 of the original Hauser report is not suitably addressed by this study, and there is a serious problem with the *internal validity* of the study. Smokers do not purchase light cigarette brands in real life in an artificial manner on a computer screen given these profile descriptions of these sole four product attributes.

10*. More importantly, this study does not show reliance or conformity with respect to smokers' beliefs when they purchase light cigarette brands. That is, the study does not, in any form, address any issues of smokers being misled about purchasing light cigarette brands, and this study's relevance to this case is questionable at best. Interestingly, a second study conducted by Dr. Hauser with an alternative sample of light cigarette smokers that was not mentioned in his initial report, does tangentially deal with this issue of smoker beliefs. Exhibit VI in my previous report displayed simple frequency distributions for the wording and responses to questions 8, 9, 10, and 11. **Almost 80% of the respondents currently believe that light cigarettes have the same or more health risks as regular cigarettes. Of these, almost two-thirds have always held that belief.** This is certainly not supportive of any reasonable claim of reliance on misleading information, as suggested by the plaintiff in this case.

11. There are several product attributes that are important to smokers (cf. Carter and Silverman, 2004) that were omitted from this survey in the conjoint analysis. And respondents in the qualitative interviews conducted prior to this conjoint study evidently mentioned many of these additional product attributes. For example, brand of cigarette was not explicitly included as an attribute in the study. This is a serious deletion in light of the fact that each light cigarette brand has its own image, and given the fact that different smokers may be attracted to different aspects of a given brand's equity, including price, packaging, image, flavor, consistency, quality, or use by family and friends. Other potentially important missing product attributes include the length of the cigarette, the type of filter, tar/nicotine content, flavor enhancements (e.g., menthol), etc. Dr. Hauser states that the use of the particular error distribution protects one against misspecification errors, as they would be accommodated by an error term. This is only true if such missing product attributes do not interact with the attributes included in the design. There is no information/evidence provided in the study that such missing attributes are insignificant or do not confound the results derived.

12. In addition, from an inspection of the model structure, no interactions were included and estimated explicitly in the model or accommodated in the design outside of a two-way interaction with price. Dr. Hauser claims that one can discern such interactions on the basis of interviews with smokers. I fully disagree with this statement as the only concrete way to examine such interactions is to utilize experimental designs that permit their estimation and actually empirically estimate and test them in the model itself. Failure to do so explicitly makes the main effects reported in this report highly suspect, as main effects are often confounded with higher order interactions and thus are

often meaningless to talk about, especially with crossover type interactions. While some diagnostics are available in the Sawtooth Software procedure to identify interactions with the factors used in the study, this feature is useless when one mis-specifies the underlying utility model and leaves out essential attributes (brand, the length of the cigarette, the type of filter, tar/nicotine content, flavor enhancements (e.g., menthol), etc.). As such, the part-worths estimated and reported, as well as the factor importances and calculations, are all suspect.

13*. In the Hauser conjoint study, respondents were presented choice sets of hypothetical light cigarettes defined with respect to specific levels of pack type, price, taste, and health risk. While the levels of the pack type are concrete and easy for all respondents to comprehend, there are potentially serious problems associated with a common understanding of the levels of the remaining three factors utilized in the conjoint analysis. For example, with respect to health risks, it is not clear how each respondent processes and interprets the different levels of health risks associated with regular, light, and ultra-light cigarettes, i.e., exactly what health risks are in the mind of each smoker at this time. Dr. Hauser's study does not accommodate differing perceptions of specific health risks. There may be no uniformity of beliefs here despite the national press on this topic. Also, there may not be uniformity with respect to respondents' perceptions of the health risks associated with these different cigarette forms across the sample. Some may believe that all cigarettes are unsafe as is now advertised heavily on television. Others may assume an implicit order of safety related to the amount of tar/nicotine in the cigarette. Because perceptions may vary, the part-worth numbers estimated per respondent may not be comparable.

14*. This same argument can be made regarding the levels tested for taste. Taste for these different cigarette forms can vary substantially depending upon the brand (brand specific additives and/or tobacco in the blend) and tar content. Furthermore, a respondent cannot reliably assume a level of taste based on assumptions of a regular or ultra-light cigarette in a computer exercise. In addition, there are individual differences and experiences with respect to perceptions and beliefs on taste. Taste is a multidimensional concept and relates to various characteristics about the cigarette related to the type of tobacco utilized, tar/nicotine content, flavor additives such as menthol, ventilation, and tobacco blend, and flavoring, as well as to individual physiological and behavioral differences. It is virtually impossible for a respondent to reliably assess the taste of any other cigarette outside of the particular brands s/he has consumed. Evidence of this perceptual ambiguity can be derived from the results of Hauser's own study. Exhibit VII in my previous document showed various descriptive statistics on comparative perceived health risks and tastes between cigarette forms from the survey. Note the huge range (0-150) associated with each of these four questions, as well as the large variances. Such large ranges and variances here support the contention of a lack of uniform or consistent perceptions with respect to health risks and taste.

15*. There are similar problems of potential perceptual ambiguity associated with price. Not everyone pays the same price for a pack of the same brand of cigarettes as there are rather dramatic differences in state associated taxes (e.g., compare the taxes on a pack of cigarettes for VA vs. NJ). Also, the price of the pack often depends on the particular brand. Again, a particular level of any one of these levels may mean different things to different smokers given their experiences. Thus, the specific percentages (50%,

20%) tested more/less than what the respondent currently pays imply different amounts to different respondents.

16*. Also, the fact that many of the conjoint questions were asked in relationship to the smoker's experience with their current brand of light cigarette is problematic. Different smokers have different experiences and beliefs about their own brands of cigarette, and these were not dealt with appropriately in this study. The benchmark smokers utilized therefore differs across the board with respect to using their current brand as a referent point. As such, it is problematic that Dr. Hauser treats the derived estimates as unconditional (comparable) over different respondents.

17. There are other potential problems associated with the various product attribute levels tested in the conjoint analysis. Researchers in the field have documented the existence of a "levels effect" in conjoint analysis where factors employing more levels tend to be higher in significance in the estimation (Steenkamp & Wittink, 1994; Verlegh & Wittink, 2002; Wittink, Krishnamurti, & Reibstein, 1989). In fact, noting that price, health, taste, and pack were tested with 5,5,3, and 2 levels respectively, it is not surprising that the order of importance also follows that same ordering. Dr. Hauser placed no controls over this "levels effect". It is therefore likely that the particular design of the experiment itself had a direct impact concerning the end result/findings with respect to the order and magnitudes of levels tested for the four product attributes.

18*. There are several problems associated with respect to the constructed hypothetical product choice sets and assumptions utilized in this conjoint analysis. For example, the smoker was told explicitly in the instructions to the conjoint analysis to assume that these four product attributes are independent, and each attribute level does not affect any other.

This is problematic in reality when it is known that taste is affected by tar content which can also impact health risk.

19*. Many of the hypothetical products are infeasible in that nothing in the marketplace even approximates such hypothetical choices. For example, it may not be possible to attain a pack of cigarettes 50% more or less than what the smoker now pays. While one often creates product choice sets in conjoint studies that stretch the actual brands in the market place in terms of attribute levels, a wide discrepancy may not be believable in practice and may provide unreliable responses. In addition, using artificially inflated levels of factors (e.g., price) may lead to having that particular factor end up becoming more important in the end result.

20*. The respondents are asked to make several assumptions that may run counter to their own perceptions or what is known in the marketplace at the time of this study. As mentioned above, respondents were asked to assume that taste and health risks were independent, which they are not. The levels of the health risk factor appear ordered or at least different, when in fact all such forms of cigarettes are not safe. The effect may be to lead the respondent and inflate the overall importance of health risks in this study. Since there was no tactile sampling of taste with actual prototype cigarettes, the smokers were told to assume various perceptual levels of taste in comparison to either their own brand of light cigarette or what they thought a regular or ultra-light brand of cigarette would be like. Serious problems occur here since different brands of such cigarettes taste different, and the perceptual benchmarks each respondent has in mind in response to this particular aspect is likely to be quite different. As such, the derived importances and part-worths

estimated here are not comparable across respondents, and this analysis should have been performed within each individual respondent separately.

21*. One of the most useful features associated with conjoint analyses is the ability to validate the workings of the model in predicting the actual market shares of existing brands in the market place. In fact, this is one of the advantages of using a CBC formulation since such share predictions come directly from the model itself as opposed to having to apply an intermediate step with a separate choice/share model. Because of the problems associated with potential mis-specification/lack of factors such as brand, cigarette length, type of filter, tar/nicotine content, etc., as well as the perceptual distortion issues raised with respect to differential individual perceptions of the various factor levels utilized in this conjoint analysis, validation of the actual brands' share in the market place is not feasible in the Hauser study. And if attempted, the results would be grossly inaccurate in many cases given the large market share differences (brand identity/equity) associated with different brands in the light cigarettes market. For example, without using brand name in such model predictions, one would be at a loss in attempting to account for share differences observed between, for example, two brands of light cigarettes that were identically priced, had hard packs, possessed the same health risks, and tasted the same, but had drastically different market shares in the market (e.g., compare Marlboro, the market share leader, with another similar generic cigarette brand). The inability to perform this *predictive validation* with real market light cigarette brands is a severe limitation and raises serious questions as to the usability of the results of this study. Using a holdout sample of choice sets for predictive testing, as Dr. Hauser did, does not provide the same confidence in the performance of the model, as does validating

real market behavior. The study misses an essential aspect of the cigarette market regarding the actual consumer perception of real market brands of light cigarettes that should have been a key aspect of any study of this form.

22. The various technical problems identified here are those where Dr. Hauser's overall research design and analysis of the resulting data do not meet accepted practice of excellence in the field of applied Statistics. First, the respondent is not allowed to be indifferent in the Hauser study and is forced to make a choice in each of these choice sets. In addition, there is not a "No Choice" option in case all options are dis-preferred, or where there is an "own-Brand" option. This does not reflect reality and can therefore bias the results. Forcing a choice for situations involving "equally dis-preferred" hypothetical brands will affect the analysis and resulting parameter estimates. The majority of CBC studies contain a No-Choice option for this reason (cf. Louviere, Hensher, Swait, and Adamowicz , 2000).

23. Hierarchical Bayes procedures have been criticized over the parametric assumptions made with respect to the prior and/or hyper prior (as well as likelihood) distributional assumptions made. As such, researchers attempt to demonstrate robustness to such assumptions by utilizing different prior assumptions and seeing how much the results change. Such sensitivity analysis has become commonplace to applied Bayesian work. Yet, such analyses were not included in the Hauser report, thus deviating from excellence in applied statistical practice. It is important to demonstrate that one's prior assumptions do not dramatically influence the results. Even if non-informative hyper-priors are utilized, these distributions follow a certain shape with characteristics about the

first four moments of the distribution (mean, variance, skewness, kurtosis) that typically affect the results obtained.

24. Most applied researchers examine residuals in their models to examine if there are structural problems observed with respect to parametric assumptions, the presence of outliers, missing explanatory variables in the model, etc. This was also not performed in the present study and there is no confirmation of the viability of the model employed here. Hauser states that the U^2 figure reported ranges from 0 to 1 as a measure of fit (the closer to 1.0, the better). The reported average fits were 0.52 for the calibration profile choices and 0.459 for the holdout profiles (p.18). Interpreting this as Dr. Hauser does as a “percentage of uncertainty explained by the model” (p.17), you have roughly half of the information in these choice sets **unexplained** by the Hauser HB model. This other 48-54% is unexplained and we cannot be assured that the calculations derived from the estimates of this model are valid when we see such fit values.

25*. The HB CBC methodology estimates individual level part-worth utilities and thus factor importances. Dr. Hauser demonstrates that the HB CBC methodology outperforms the aggregate MNL conjoint model that assumes one common set of estimates for the entire sample. These attribute importances for the four conjoint product attributes from the HB CBC analysis were plotted in Exhibit VIII of my previous report via box plots to demonstrate an important point to be made. As shown in that figure, the interquartile ranges for the taste and health risk factors overlap substantially. For a sizable percentage of respondents, taste is actually more important than health risk. This heterogeneity amongst respondents means that not all respondents are alike and that there is no typicality or uniform thought present in this marketplace.

26*. Hierarchical Bayesian methods offer the advantage of deriving individual level estimates of model parameters with relatively sparse data. This procedure pools over all the respondents in the various calculations involved in the MCMC algorithm. However, as mentioned earlier, if the data is not truly comparable over all respondents, then there is a problem with the application of this procedure. In particular, if the respondents are each utilizing different perceptual beliefs and benchmarks (e.g., responding within their own brand) in interpreting the various conjoint factor levels provided in the various choice sets, then the analysis needs to be conducted on an individual by individual level – an aspect that cannot be accommodated with the choice set designs utilized in this study (especially incorporating a saturated interaction model).

II. HAUSER'S RESPONSE

Hauser's written response to the court dated 7/14/06 was intended to address the many issues levied against his initial study by all of the defense team's experts. In carefully reviewing that document, the majority of his text deals with criticisms levied by the defense team experts other than myself. In fact, the majority of the 26 issues I raised initially were not responded to. The fact that several issues remain unchallenged seriously questions the validity of the study as indicated above and in my original submission. I will therefore limit my discussion to those responses of Hauser that deal with the particular issues from my previous submission. I will summarize the Hauser's response to the issues raised and refer to the section and page numbers in Hauser's response. I will then respond in bold italics with my rebuttal.

In section V of his recently submitted report, Hauser claims in pp.12-22 that the sample he selected is representative of the smokers of light cigarettes in the US:

(a). Hauser implicitly acknowledges that the sample quotas he initially utilized may not be accurate and re-weights the data according to the other distributions provided in the experts' reports according to several demographic variables he utilized to establish quotas to draw the sample of 627 Greenfield opt-in panel members (plus ethnicity and education). He shows in his Table 3 that two specific results (not all the results were shown) do not change from the un-weighted analysis, and therefore concludes that the sample is representative.

Unfortunately, this analysis does not prove representativeness of the U.S. population of smokers of light cigarettes. As mentioned in my original critique summarized above, it is estimated that approximately 25% of the U.S. population does not utilize the Internet. I do not see any representation from this segment of the population in this sample. Also, I do not see any representation concerning the vast majority of smokers of light cigarettes that do not participate in this Greenfield panel. Similarly, there is no representation regarding the 84% of the Greenfield panel who refused to participate. Hauser does not prove that their sample is representative of these other major groups of smokers of light cigarettes with respect to the issues being studied involving utility preference functions, trade-offs, and choice decisions (not demographics).

(b). Hauser's stated justification for balancing the sample on the four demographic variables was based on the convenient number of categories resulting (72) and the ease to

fill the quotas. Had other demographic categories such as race, marital status, occupation, and employment status been added, then the total number of categories would have ballooned out of proportion with some 15,552 cells which is much larger than the sample size of 627. He states “the expert makes a decision about which categories to use in balancing the sample. This is a common practice and is used extensively in the industry” (p.17).

The use of any form of stratification or quota variables assumes that the criteria utilized to form the sampling cells are strongly related to what is being investigated, i.e., the use of these four demographic variables utilized for quota sampling tacitly assumes that these variables are highly related to these underlying attribute importances, utility part-worth functions, and resulting choice behaviors. As shown in my previous analyses in Exhibit III in my previous report, there is no strong argument, either theoretical or empirical, to support this conjecture. These demographics were selected on the basis of convenience and there is no theoretical linkage to the underlying behavior phenomena being studied. One could just as well have utilized eye or hair color as a basis for setting quotas. In fact, this conclusion is very much in line with the research performed by Moore (1980) which documented the lack of relationship between individual differences in conjoint analysis part-worths and such demographic variables.

(c). Hauser attempts to counter my criticism about the use of professional survey takers by performing an analysis with heavy vs. light survey respondents and showing that two

results are somewhat similar, although noting a tendency for the lighter survey group to value health risk more.

In order to fully dismiss my criticism, comparisons have to be made with smokers of light cigarettes who are not members of this panel (professional survey takers) that was not done. And, comparisons need to be made with the entirety of the results, not just two aspects of the analysis as he has done here and above with the demographically weighted analysis. Thus, Hauser has not proven that the Greenfield panel survey of 627 members is truly representative of the US smoker of light cigarettes population with respect to their underlying preferences, utility functions, choices, and trade-offs. At best, he may have a demographically representative sample of smokers of light cigarettes who have Internet access, were members of the Greenfield panel, and who chose to participate in this particular conjoint task.

One has to move to section XV of his responses (starting on p.54) to obtain any further Hauser response to my set of criticisms levied in my May 2006 submission. On pages 54-56 of Hauser's responses, he defends his not using brand as an explicit feature in the conjoint study, stating that the goal of the task was to obtain choices within the respondent's own brand. He states, in support of this decision, that there were no significant differences found by brand name for health risk.

First, in re-examining the instructions to the CBC portion of the questionnaire, I did not see explicit instructions for the respondent to make these particular choice decisions within his/her own brand of light cigarette (see the respondent instructions to the choice task on E13 screen capture in Hauser's earlier report). (However, I do not

have access to the actual screen captures of the choice task itself that presents the choice sets where such instructions may have been added.) Secondly, there is a statistical issue concerning sufficient degrees of freedom to test brand differences across all brands given the sparse sampling in many of the brands of light cigarettes collected, and this is why one has to either pool over these “sparse” brands or restrict the analysis to only the top set of brands for which substantial representation in the sample occurs. Third, I take issue with the implications of Hauser’s assessment concerning differences by brand. I performed ANOVA’s using brand as the independent factor (11 levels including the top 10 brands plus pooling over all others) and the various part-worths estimates by respondent. Below are the resulting ANOVA tables and they show definite significant ($p < .10$) brands effects with respect to selected part-worths across three of the four factors included in the study.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Partworth for pack type: soft pack	Between Groups	3.929	10	.393	2.136	.020
	Within Groups	113.319	616	.184		
	Total	117.248	626			
Partworth for pack type: hard pack	Between Groups	3.929	10	.393	2.136	.020
	Within Groups	113.319	616	.184		
	Total	117.248	626			
Partworth for health risk: gt than a regular cigarette	Between Groups	25.899	10	2.590	.816	.613
	Within Groups	1955.466	616	3.174		
	Total	1981.365	626			
Partworth for health risk: eq to a regular cigarette	Between Groups	2.703	10	.270	.659	.762
	Within Groups	252.487	616	.410		
	Total	255.190	626			
Partworth for health risk: eq to a light cigarette	Between Groups	3.033	10	.303	.954	.483
	Within Groups	195.925	616	.318		
	Total	198.958	626			
Partworth for health risk: eq to an ultra-light cigarette	Between Groups	6.269	10	.627	.979	.460
	Within Groups	394.333	616	.640		
	Total	400.602	626			

Partworth for health risk: It an ultra-light cigarette	Between Groups	9.049	10	.905	.698	.727
	Within Groups	798.258	616	1.296		
	Total	807.306	626			
Partworth for taste: like a regular cigarette	Between Groups	15.651	10	1.565	2.034	.028
	Within Groups	474.062	616	.770		
	Total	489.712	626			
Partworth for taste: like a light cigarette	Between Groups	11.842	10	1.184	1.167	.310
	Within Groups	625.271	616	1.015		
	Total	637.113	626			
Partworth for taste: like an ultra-light cigarette	Between Groups	25.400	10	2.540	1.763	.064
	Within Groups	887.674	616	1.441		
	Total	913.074	626			
Partworth for price: 50% less than usual	Between Groups	100.457	10	10.046	2.735	.003
	Within Groups	2262.556	616	3.673		
	Total	2363.013	626			
Partworth for price: 20% less than usual	Between Groups	16.190	10	1.619	1.406	.173
	Within Groups	709.318	616	1.151		
	Total	725.508	626			
Partworth for price: same as usual	Between Groups	5.584	10	.558	.546	.857
	Within Groups	629.670	616	1.022		
	Total	635.254	626			
Partworth for price: 20% more than usual	Between Groups	31.251	10	3.125	2.365	.009
	Within Groups	813.918	616	1.321		
	Total	845.170	626			
Partworth for price: 50% more than usual	Between Groups	96.880	10	9.688	1.666	.085
	Within Groups	3583.154	616	5.817		
	Total	3680.034	626			
Partworth for pack type: soft pack	Between Groups	.029	10	.003	.896	.537
	Within Groups	1.973	616	.003		
	Total	2.002	626			

In fact, these differences are still significant when using robust means tests as shown

below:

Robust Tests of Equality of Means

		Statistic(a)	df1	df2	Sig.
Partworth for pack type: soft pack	Welch	2.250	10	84.409	.022
	Brown-Forsythe	2.073	10	91.916	.035
Partworth for pack type:	Welch	2.250	10	84.409	.022

hard pack	Brown-Forsythe	2.073	10	91.916	.035
Partworth for health risk: gt than a regular cigarette	Welch	.832	10	83.860	.599
	Brown-Forsythe	.823	10	216.123	.607
Partworth for health risk: eq to a regular cigarette	Welch	.622	10	84.313	.792
	Brown-Forsythe	.692	10	226.147	.732
Partworth for health risk: eq to a light cigarette	Welch	.962	10	83.977	.482
	Brown-Forsythe	.988	10	220.543	.455
Partworth for health risk: eq to an ultra-light cigarette	Welch	1.050	10	84.032	.410
	Brown-Forsythe	1.010	10	215.433	.436
Partworth for health risk: lt an ultra-light cigarette	Welch	.651	10	83.397	.766
	Brown-Forsythe	.664	10	207.083	.757
Partworth for taste: like a regular cigarette	Welch	1.980	10	83.512	.046
	Brown-Forsythe	1.959	10	185.547	.040
Partworth for taste: like a light cigarette	Welch	1.200	10	83.553	.303
	Brown-Forsythe	1.117	10	172.020	.352
Partworth for taste: like an ultra-light cigarette	Welch	2.799	10	84.986	.005
	Brown-Forsythe	1.852	10	225.621	.053
Partworth for price: 50% less than usual	Welch	2.845	10	84.010	.004
	Brown-Forsythe	2.804	10	183.527	.003
Partworth for price: 20% less than usual	Welch	1.514	10	83.733	.149
	Brown-Forsythe	1.410	10	194.042	.178
Partworth for price: same as usual	Welch	.493	10	83.692	.890
	Brown-Forsythe	.534	10	185.244	.864
Partworth for price: 20% more than usual	Welch	2.409	10	83.985	.014
	Brown-Forsythe	2.426	10	196.115	.010
Partworth for price: 50% more than usual	Welch	1.739	10	83.658	.085
	Brown-Forsythe	1.645	10	193.275	.097
Partworth for pack type: soft pack	Welch	.841	10	83.462	.590
	Brown-Forsythe	.858	10	190.903	.573

a. Asymptotically F distributed.

Hauser seems to “cherry-pick” the results he chooses to report concerning no significant differences in health risk. However, there are significant differences by brand with respect to pack type, taste, and price. I find this lack of full disclosure of the results very disconcerting especially since these results do not support Hauser’s non-use of brand name as an explicit conjoint attribute. The fact that there are significant differences for the taste and price factors suggest that conclusions concerning the Health vs. Taste comparisons, as well as all the willingness to pay

calculations, must be conditioned by brand, which was not done. As a result, these computations are all suspect.

In Section XVI of Hauser's latest response, he states that the four features/attributes were selected "appropriately" to address the issues studied. He claims that no other attributes were deemed necessary to study the tradeoff between health risk and price, and that no interactions were found to be significant to include on the basis of the qualitative interviews, pretest interviews, and the Sawtooth CBC Software Counts heuristic.

The fact is that actual consumers in qualitative interviews conducted by AMS did mention other attributes in the initial preliminary interviews as important determinants of light cigarette purchase/choice decisions in the actual marketplace, and these other attributes were seemingly ignored in this study. Hauser's exclusion of important attributes such as brand name, the length of the cigarette, the type of filter, tar/nicotine content, flavor enhancements (e.g., menthol), etc. that the smokers of light cigarettes find important in the market choices they make as cited in my earlier report is very problematic in the present context. As mentioned in my original report, it can lead to model mis-specification. As a result, Hauser chose to perform a nonsensical academic exercise that bears little resemblance to how smokers purchase light cigarettes in the real world. In addition, without the inclusion of such important attributes, none of the purported tests of interactions Hauser claims to have done would be relevant since these attributes were not explicitly included in the model itself. For example, Hauser provides no proof that a brand name/health risk/menthol three-

way interaction is not a significant factor. It is virtually impossible to test for such interactions in qualitative interviews. And, the Counts procedure cannot possibly detect such interactions involving attributes not explicitly included in the study. The potential for model mis-specification can explain the fact that nearly half of the uncertainty remains still unexplained here.

Hauser attempts to answer my criticism over potential “levels effects” (different attributes are expressed in terms of different numbers of attribute levels in his study) on p. 58 and A10 in Appendix B, and argues that his major focus was on the health risk and price trade-offs where both attributes were measured in his study at the same number of levels (5). Thus, the levels effect is not pertinent to his conclusions.

As I have stated in my initial summary at the outset of this report in Section I, the significance of the four factors follow the same order as their number of levels. That is, factors specified at more levels are more significant than factors specified at fewer levels. Hauser fails to report in his rebuttal that taste was only measured at four levels in his design, and that a good portion of his initial report is dedicated to attempting to illustrate that health risks are overall more important than taste in driving these choice decisions. As such, one does not know whether this result is an artifact of the experiment due to the ad hoc manner in which these attribute levels were specified or whether consumers actually value health risk over taste. Needless to say, the design of any research should not impact the findings, and no proof of the opposite has been provided in the various Hauser documents or attempted response.

Hauser argues on p.60 and in A14-A15 that his U^2 statistic, the percentage of uncertainty that his model explains, are “well within the ranges accepted in peer-reviewed scientific publications in Marketing Science” (p. A14).

My original comment was that the reported average fits were 0.52 for the calibration profile choices and 0.459 for the holdout profiles. Interpreting this as Dr. Hauser does as a “percentage of uncertainty explained by the model”, you have roughly half of the information in these choice sets unexplained by the Hauser HB model. Hauser lacks any explanation as to what causes this other 48-54% unexplained variation. As such, there is no assurance that the calculations derived from the estimates of this model are valid when we see roughly half of what is going on in the data unaccounted for. No explanation at exactly what is accounting for the other 50% of the variation is provided. It may be model mis-specification (e.g., other attributes and interactions were not included in the model), respondent error (e.g., they did not take the task seriously), violated model assumptions (e.g., the particular distributions used), etc. Hauser offers no attempt at explaining this other half of the coin and wants us to blindly accept his perception of academic standards. Note, to my recollection, I have not written or published a conjoint related paper using this U^2 measure.

Hauser’s comment related to my initial criticism concerns the no choice and forced choice option utilized in his conjoint experiment. His response on p.58 and A8-A9 states that the forced choice procedure has a long history in marketing, has been applied for years, and is well-accepted practice. He is not aware of any studies conducted

“that show the inclusion of a no-choice option changes any of the estimates in a substantial way” (p. A9).

I wish to bring attention to some recently published literature in this area which contradicts Hauser and discusses the types of problems that can arise by using forced choice data collection in such studies. A forthcoming article in Marketing Letters by Brazell et al (2006) states the following:

“It has long been advanced that one should include a constant or no-choice option in choice-based conjoint designs. This alternative can be an option, such as ‘keep on shopping’ that is the same for all respondents, or it can be an option like ‘stay with my current product’ that would vary across, but is constant within respondent. For ease of exposition, we will refer to this option as the ‘no-choice option’. Inclusion of the no-choice option increases design efficiency, better mimics the choice process in many situations directly measuring demand for specific tested products in the context of the entire market, and allows one to model market growth as more attractive alternatives are introduced.” (p.256).

These authors provide an alternative response format involving dual response given the loss of information provided with a huge amount of no-choice responses.

Haaijer, Kamakura, and Wedel (2001,2002) have also recognized the importance of including this no-choice option in CBC designs in representing real world purchase decisions:

“In order to make the choice more realistic, in many conjoint experiments one of the alternatives in the choice sets is a ‘no-choice’ or ‘none’ option. This

option can entail a real no-choice alternative (none of the above) or an ‘own choice’ alternative (I keep my own product)” (p.93 of the 2001 article). Haaijer et al. (2001) propose a variety of efficient ways of coding such ‘no-choice’ options in such CBC designs.

On the more behavioral and theoretical side, Dhar (1997) and Dhar and Simonson (2003) have experimentally demonstrated the effects of forced choice data collection on actual consumers’ utility functions, as well as having a no-choice option.

Finally, Hauser responds to my criticism of his not performing a sensitivity analysis with different prior distributions in examining their effect on the results by stating he used non-informative priors, and that I published papers in this area not following my own advice.

It should be noted for the record that the criteria for publishing articles in the top methodological journals that I publish in are very different than the criteria utilized to assess applied work such as what Hauser has done. In the A level journals I publish in, the emphasis is on developing a new methodological procedure that either accomplishes something that no other procedure in existence does, or performs better than existing procedures. The emphasis is on methodological rigor and nuance, and not on applications per se. An application is usually provided to illustrate the workings of the procedure in contrast with competing techniques, with appropriate comparisons made. Needless to say, the present study conducted by Hauser would be considered an application of existing procedures and “old-hat”, and would not be acceptable in any A level journal of this caliber. In addition, the vast majority of the journals I publish in

have a page limit (around 32-40 pages) and it is impossible to examine every detail of a proposed new procedure in such a manuscript, especially when responding to Reviewers' comments. However, when one is dealing with only an application of existing technology, different questions are asked as to the applied nature of the problem at hand. For example, with an application, one is often concerned about whether the model assumptions hold and how violations affect the results. Thus, the criterion for excellence is quite different than in published articles in A-level methodological journals, and Hauser is well aware of this fact as a previous journal editor.

As I mentioned in my previous response, the use of non-informative hyper prior distributions are not a protection against model mis-specification. Rather than provide a tutorial summarizing decades of statistical research on the need for such sensitivity and model checking analysis in such hierarchical Bayes models, I refer the reader to Chapter 6 of the book entitled Bayesian Data Analysis by Gelman, Carlin, Stern, & Rubin (1995, 2003) which talks about the problems that result from improperly specified priors and likelihoods in such applied work, as well as a variety of recommended diagnostics and analyses- none of which were reported in the original Hauser document.

III. MY SUMMARY ASSESSMENT

My initial professional assessment of the Hauser study was that it was fraught with a variety of problems discussed above (some 26 different sets of problems) and that the results are virtually meaningless in the context of describing the actual

decision/choice processes engaged by light cigarette brand smokers. The conjoint task is unrealistic, not reflective of how actual smokers make decisions concerning light cigarette brands, and conducted in an artificial setting. The real world decisions faced by smokers of light cigarette brands in the purchase and consumption of light cigarette brands have not been sufficiently investigated in this study. Important product attributes deemed essential to various decisions concerning light cigarette brands have been ignored. The sample selected for use is biased, non-representative, and the results of this survey study cannot be generalized to the population of light cigarette brand smokers within prescribed confidence and precision limits. Serious experimental problems potentially exist with respect to a uniform understanding by respondents of the levels of the product attributes tested, as well as the number and extreme ranges of the levels of some of the attributes employed in the study that could seriously affect the results. As such, the external and internal validity of this work is highly questionable. Modeling assumptions have not been adequately tested and verified, nor has proper sensitivity analysis of the results to different assumptions been performed. There has not been any appropriate addressing of issues concerning reliance, conformity, or causation in this study, especially given the massive amount of heterogeneity amongst these smokers. What little work there has been on reliance/conformance occurs in Study #2 which is not even mentioned in the Hauser report. After my analysis of Hauser's response, my initial assessment remains the same.

Wayne S. DeSarbo
Dr. W.S. DeSarbo

8/25/06
Date

REFERENCES

Brazell, J.D., C. Diener, E. Karniouchina, W. Moore, V. Severin, & P. Uldry (2006), Marketing Letters, 17, 255-268 forthcoming.

Dhar, R. (1997), "Consumer Preference for a No-Choice Option", Journal of Consumer Research, 24, 215-231.

Dhar, R. & I. Simonson (2003), "The Effect of Forced Choice on Choice", Journal of Marketing Research, 40, 146-160.

Haaiker, R. W. Kamakura, & M. Wedel (2001), "The No-Choice Alternative in Conjoint Choice Experiments", International Journal of Market Research, 43, 93-106.

Haaiker, R. W. Kamakura, & M. Wedel (2002), "Alternative Specifications to Account for the No-Choice Alternative in Conjoint Choice Experiments", 2001 Sawtooth Software Conference Proceedings, Sequim, WA.

Gelman, A., J. Carlin, H. Stern, & D. Rubin (1995), Bayesian Data Analysis, Chapman & Hall: London, UK.

Gelman, A., J. Carlin, H. Stern, & D. Rubin (2003), Bayesian Data Analysis, 2nd edition, Chapman & Hall: London, UK.

Moore, W.L. (1980), "Levels of Aggregation in Conjoint Analysis: An Empirical Comparison", Journal of Marketing Research, 17, 516-523.