

Discussion

Ideographic sensory testing vs. nomothetic sensory research for marketing guidance: comments on Garber et al.

Armand V. Cardello*

US Army Natick Soldier Center, Natick, MA 01760-5020, USA

Garber et al.'s paper joins the growing chorus of calls for changes in the way that sensory tests are conducted. The primary issue raised by the authors of these appeals concerns the sterile nature of sensory testing. They point out that sensory researchers usually examine one or a small number of independent variables for their effect on a single dependent measure of sensory experience, while taking extraordinary precautions to rule out confounding variables through the use of strict controls on environmental, contextual, and situational variables. The essence of their argument is that the data obtained from such sterile tests do not predict consumer behavior toward foods in the real world. This view has been colloquialized in the call to test "real foods, with real people in real situations" (Meiselman, 1992).

In this latest overture, Garber et al. focus their argument on the goal of assisting marketing researchers to more "accurately predict consumer behavior at the point of purchase". The authors enumerate a number of recommendations and strategies for improving sensory test designs to better predict the food preferences and choices of consumers in the marketplace. Although their recommendations are logically consistent with the goals of marketing research, they are presented somewhat too blithely and without due consideration of their implications for the advancement of the goals of sensory science. If adopted unquestioningly by practicing sensory analysts, the recommendations would subjugate future progress in sensory science to the exigencies of marketing research. In the rush to restructure how sensory data are collected to improve prediction of marketplace behavior, risks would be posed to one of the fundamental goals of sensory science—the development of explanatory theories of the role of sensory functioning in food behavior.

I'll begin by raising some logical counterpoints to the recommendations made by Garber et al. and then

address the more general philosophical issues and implications of adopting their recommendations without due consideration to the consequences for sensory science.

At the outset, I am pleased that the authors have acknowledged the differences between "theory-oriented" and "applications-oriented" research and have stated that external validity may be more important to the latter form of activity. This point is noteworthy, because there is a sharp dichotomy in the sensory sciences between *practitioners*, who are product-focused and involved with solving day-to-day problems of sensory analysis, and *basic researchers* who are focused on identifying causal mechanisms and general principles of sensory functioning in food-related behavior. Although they differ in their goals, there is a kindred dialectic between basic and applied researchers in the field. On the one hand, basic researchers utilize the published empirical data obtained by sensory practitioners and other food scientists as scientific fodder for developing viable hypotheses to explain consumer food behavior and for post-hoc hypothesis testing. In turn, practitioners use the theories derived from basic sensory research to aid in the design of their applied studies and to interpret the meaning of their empirical data for customers. Into this dialectic of theory-oriented and applications-oriented sensory research, Garber et al. enter their petition to increase the external validity of sensory research by incorporating specific subject samples (target populations), product ranges (consideration sets and product parities), scalar methods (unvalenced), and informational variables (the marketing mix) into sensory test designs. Although these recommendations are well intended and consistent with the goals of predicting marketplace behavior, in their extreme they tilt the dialectic between sensory theory and application toward applied, ideographic research and away from the nomothetic research imperative that is the fundamental basis of meaningful scientific inquiry.

Let's begin by looking at the authors' first recommendation, i.e. that sensory testing should utilize consumer segments that represent the consuming population of the product. Although this recommendation would certainly improve the prediction of responses of consumers who are members of the current target population, it would not be useful for predicting the responses of consumers who are *not* in the target base, i.e. consumers who are new to the food category. Neither would it be useful for uncovering the sensory and perceptual mechanisms that underlie *avoidance* of the category. To the extent that sensory researchers limit test populations to specific subgroups in the population, the resulting data cannot be generalized beyond those subgroups. Since one of the main objectives of sensory science is to uncover the basic visual, gustatory, olfactory, and tactile mechanisms that influence approach and avoidance behavior toward foods, restricting test populations to only current consumers of the product will preclude uncovering the very factors responsible for the evolution of the product's "target population" in the first place.

The authors' second recommendation is that sensory research should restrict the contextual range of products to a small subset of similar alternatives, because most consumers in the marketplace choose from among a small "consideration set" of products, the members of which all have approximate "parity". Here again, the recommendation is designed to improve prediction of choice in a specific market setting; but what happens if the market setting changes? What if new product introductions alter the consideration set? Are the data still valid? The effect of stimulus range in sensory tests has been well studied and has resulted in a robust literature on perceptual contrast effects and contextual dependencies in sensory testing. The data are indisputable in showing that both sensory and affective responses are extremely sensitive to the nature and range of contextual stimuli. Thus, any change in the consideration set as the consumer moves from one marketplace setting (e.g. supermarket, specialty grocer, warehouse store, etc.) to the next will change the predicted choice behavior. In addition, by restricting the stimulus range to products that have parity (i.e. similar sensory or hedonic quality), the probability of identifying differences between the test sample and the consideration set will be maximized. While this might be considered advantageous, it introduces a significant testing bias into the data, because the product differences observed under a restricted stimulus range may not be discerned when the same products are framed within a broader stimulus range, e.g. when a new market item is added to the consideration set.

Now, the devil's advocate might say that, except for monadic tests, every sensory test must have a stimulus context, so why not choose a set of comparative stimuli

that is appropriate to the end use of the data? To this, I would argue that the whole proposal to mimic the consideration set of market products is logically misdirected, because consumers never taste products in a comparative context in the real world. Only choice and purchase behavior are made in situations where one product is directly compared to another (of the same type). Real world sensory testing, i.e. tasting and consumption during meals, is always made in a monadic fashion relative to any one item. Any context that exists is created by the *other* items in the meal, which are not part of the consideration set to which Garber et al. refer. Thus, the recommendation to utilize a range of similar marketplace products in laboratory sensory tests actually creates an artificial stimulus context that never occurs in the real world.

As a consequence of the recommendation to test products within a small consideration set that includes only "acceptable" commercial products, the authors also recommend the use of a scale that is only positively valenced. This raises two issues. First is the question of whether the use of a scale with only positive valence will affect the sensitivity and reliability of the data that are collected. Oddly, Garber et al. never consider statistical criteria. Rather, their argument rests simply on the fact that unacceptable products are not part of the consideration set, so negative valence need not be part of the measurement scale. In point of fact, there is no evidence that switching to a positively valenced scale would improve the statistical properties of the data that are collected. Although the available research is limited to magnitude estimation scaling, the data suggest that either there is no difference in precision, sensitivity or reliability between unipolar and bipolar hedonic scales (Pearce, Korth, & Warren, 1986) or that bipolar hedonic scales have greater discriminability than unipolar scales (Moskowitz & Fishken, 1979). Frankly, if one really wants to improve the measurement properties of the scales used in sensory testing, category scales (regardless of valence) should not be used. Rather, researchers should adopt ratio scales or labeled magnitude scales of liking. The latter type of scale enables ratio level judgments, improved sensitivity, and equal reliability to the 9-point hedonic scale, while maintaining similar ease of use (Schutz & Cardello, 2001). Such advantages could well justify a change from the 9-point hedonic scale. However, without statistical or other criteria as a basis, Garber et al.'s recommendation to convert to an unvalenced scale is unwarranted.

The second issue raised by this recommendation concerns the importance to any science of having a single benchmark measurement technique. Prior to the development of the 9-point hedonic scale, sensory researchers employed a wide variety of different scalar methods. The multiplicity of methods made comparisons of data between studies almost impossible. With the develop-

ment of the 9-point hedonic scale in the 1950s, the majority of sensory researchers in the food area adopted this scale as their primary measurement tool. The 9-point hedonic scale has now gained widespread, international use in the food industry. The use of a common scale among researchers has the advantage of enabling cross-comparisons of data. Converting to a different scale without some demonstrated statistical or theoretical justification is unwise, since it undermines the ability to compare results between studies and research laboratories.

The last recommendation made by the authors is to incorporate the “marketing mix” into routine sensory tests, either through instructions, background information, and test conditions that mimic market factors or through the use of multiattribute measurement. Certainly, post-hoc attempts to assess the marketing mix using multiattribute analysis are welcome. However, establishing the marketing mix through information and other aspects of the test conditions will necessarily introduce cognitive, situational, and other confounding variables into the data. As someone who began his career examining the responses of single taste papillae on the human tongue, but who now studies higher-level influences on food acceptance, I have no objection to introducing cognitive or situational variables into sensory test designs, *if* they are a part of a sound, theoretically oriented research strategy (see below). However, introduction of such variables into routine sensory tests can only confound the interpretation of the results and make it impossible to generalize the data to other test conditions in which even a single variable is altered.

Taken together, the purpose and effect of Garber et al.’s recommendations are to improve the predictability of consumer responses in specific market contexts. To the extent that this is the agreed upon goal of sensory research, the authors’ recommendations are a reasoned approach to achieve this end. However, if one believes that the goals of sensory science are, *instead*, to achieve understanding of basic sensory mechanisms and to develop explanatory principles of how sensory experience influences food behavior, then these recommendations are counterproductive to scientific progress in the field.

As noted in the introduction, there is a fundamental difference between ideographic research, in which the goal is to understand and predict a *single case*, and nomothetic research, in which the goal is to establish general laws and theories that can predict *all cases*. Garber et al.’s approach would shift sensory science from a nomothetic enterprise that selects dependent variables on the basis of general principles, likely causal mechanisms, and theories of how these variables influence food product acceptance, to an ideographic enterprise that selects test conditions on the basis of factors unrelated to theory, but on unique event states of inter-

est. While the latter approach may produce *empirical* facts and predictive associations, it will not facilitate the development of *theoretical* laws and causal principles in sensory science.

The philosopher of science, Phillip Kitcher (1989) has written on the importance of theory development and unification in science. In Kitcher’s analysis of scientific explanation, he uses the hypothetical abstraction of “total science” to refer to the state of science at the end of scientific inquiry, when all facts about the universe are known. At this futuristic and logical end-point of science, billions of empirical facts about the universe will be known. Kitcher argues that the ultimate success of the scientific enterprise will be measured *not* in the number of individual facts that are known, but in the degree to which these facts have been systematized into a small, coherent, and unified set of explanatory schema. In order to achieve this ultimate goal, scientific inquiry must be continuously focused on developing unifying theories of explanation and not merely on the accumulation of specific facts that pertain to single events.

So, is there a solution to the problem of improving the prediction of marketplace behavior from sensory testing? Is there a compromise between maintaining the status quo in sensory science versus fixed adoption of the marketing-oriented recommendations proposed by Garber et al.? I think there is, and hints of the solution appear in the last few sentences of Garber et al.’s paper. For it is only here that the authors suggest that sensory scientists “manipulate” variables and examine “main and interaction effects”. Herein lies the heart of the solution for bridging the gap between sensory science and marketing research. It is the development of a joint program of nomothetic research in which marketing researchers, basic sensory scientists, and sensory practitioners all join in a collaborative effort of *systematic, exploratory* research on the effects of market-related variables on sensory responses. By bringing together the empirical power inherent in both large-scale marketing studies and applied sensory tests with the basic research goals of *theory generation* and *hypothesis testing*, significant progress can be made in uncovering the causal mechanisms by which market variables influence sensory responses and consumer behavior. Only by joining in such a concerted effort of parametric research can we develop a meaningful empirical base from which general hypotheses and principles can be drawn and tested. Using such an approach, we will be able to support Garber et al.’s objectives of improving prediction of consumer responses in the marketplace. However, the critical difference is that we will have achieved this mutually beneficial objective, while advancing what should be the primary goals of *both* sensory science and marketing research—the development of a unified set of causal mechanisms and explanatory theories to account for consumer food behavior.

References

- Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher, & W. Salmon (Eds.), *Scientific explanation, Minnesota studies in the philosophy of science* (Vol. XIII) (pp. 410–505). Minneapolis: Minneapolis University Press.
- Meiselman, H. L. (1992). Methodology and theory in human eating research. *Appetite, 19*, 49–55.
- Moskowitz, H. R., & Fishken, D. (1979). What effects do repeated evaluations play on fragrance perceptions. *Perfumer and Flavorist, 4*(May), 45–52.
- Pearce, J. H., Korth, B., & Warren, C. B. (1986). Evaluation of three scaling methods for hedonics. *Journal of Sensory Studies, 1*, 27–46.
- Schutz, H. G., & Cardello, A. V. (2001). A labeled magnitude (LAM) scale for assessing food liking/disliking. *Journal of Sensory Studies, 16*, 117–159.