The Institute for Perception

For technical reports from The Institute for Perception or for information about short courses, please visit www.ifpress.com or email us at mail@ifpress.com

Reprinted from IFPress (2013) 16(1) 3,4

When Are Two Products Close Enough to be Equivalent?

Benoît Rousseau and Daniel M. Ennis

Background: A persistent dilemma in comparing products is to know when the difference between them becomes consumer relevant. Since the probability that any two products are exactly the same is zero[1], rejecting the null hypothesis that they are identical may not be that informative; an analysis will always result in such rejection provided there is a sufficient sample size[2]. What is more important to know is whether the products differ by enough for their difference to be consumer relevant[3] - establishing this knowledge is an outstanding problem in the sensory and consumer research fields and various methods have been suggested. One method is to benchmark product differences from past tests where it is known that consumers continued to purchase the product even in the presence of variability or change. This method can be employed when products are routinely made in different factories or when blend and flavor modifications have already been introduced without any appreciable loss of sales. Another approach is to link internal panel measurements with consumer hedonic response to the set of differences[4]. But ideally, it would be useful to have a consumer-based estimate of the average criterion that consumers use to decide whether products are the same or different. In this report we discuss a way of developing this method through the use of same/different judgments.

Scenario: Your company produces a variety of condiments that are used when preparing convenience food such as sandwiches. Your responsibilities involve the qualification of alternative ingredients submitted by your suppliers, as well as the investigation of raw materials that can provide benefits in terms of lower cost or a simplified manufacturing process. You are aware that any ingredient change will result in some level of sensory difference between your gold standard products and their alternatives. You are interested in quantifying the size of that difference and in determining whether consumers would consider a change of that size acceptable. You would like ultimately to develop an action standard for each product category, but business considerations encourage you to focus first on your full fat mayonnaise line of products.

You initially consider conducting research that would link the sensory difference measured with your internal panel to preference expressed by selected groups of regular users of full fat mayonnaise[5]. However, this approach requires conducting a series of investigations with the need for the size of the underlying differences to be fairly well spread across the difference continuum. A more direct alternative is to use the same-different method, which contains information about the maximal difference that will still elicit, on average, a "same" response from a selected group of consumers. This information can be inferred from the responses collected from the same and different pair presentations.

		Response	
		"Same"	"Different"
Pair Presented	Identical	"Same"/Identical	"Different"/Identical
	Different	"Same"/Different	"Different"/Different

Table 1. Four possible types of responses from a same-different test.

Modeling the Same-Different Method: The same-different method involves the presentation of pairs of items, sometimes putatively identical, sometimes different. The task of the subject is to indicate whether he/she thinks that they are the same or different. The data is then recorded as shown in Table 1. A standard statistical analysis consists of performing a χ^2 test when all measurements are independent, or a McNemar's test when replicated measurements are involved. However, this analysis will only provide information on whether the two products are statistically significantly different, which is highly dependent on the sample size used in the experiment, and will not provide information on the size of the sensory difference.

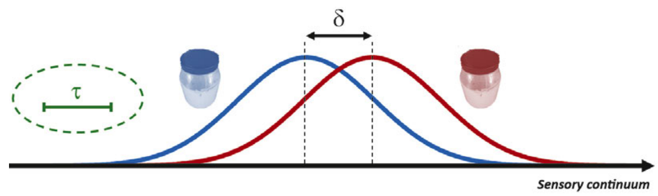

Figure 1. Thurstonian representation of product differences and τ criterion.

An alternative is to use a Thurstonian model for the same-different task. Now we can estimate the size of the underlying sensory difference between the products as measured by δ. See Figure 1. Models to estimate δ have been developed for a multitude of protocols, including the triangle, 2-alternative forced choice (2-AFC) and tetrad tests. However, the same-different method has this particularity that it also allows for the estimation of an additional parameter[6,7,8] namely the decision criterion τ. Like δ, τ is measured in terms of the standard deviation of the products' underlying perceptual distributions. When a subject evaluates two samples and the distance between the two momentary perceptions is smaller than τ, the subject answers "same" (Figure 2, Trial 1). If the distance is larger than τ, the subject answers "different" (Figure 2, Trial 2).

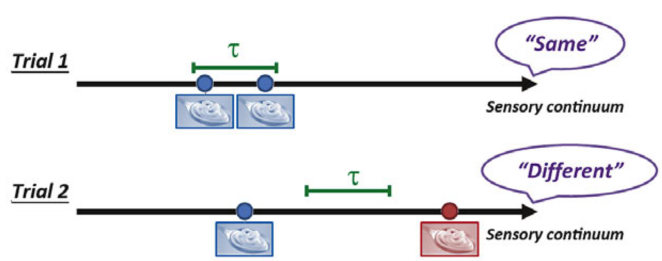

Figure 2. Illustration of "same" and "different" responses based on momentary sample perceptions.

Reprinted from IFPress (2013) 16(1) 3,4

Data Collection and Estimation of τ: In order to investigate the size of τ for the consumers of the full fat version of your mayonnaise product, you plan a study involving four different products, including your current product, with 300 consumers. While mayonnaise is seldom tasted on its own, a conservative approach is to evaluate it neat. With four products, a total of ten pairs can be presented. Each respondent receives one of each of the possible pairs, over a two day period (five pairs per day), in a randomized and balanced design. For each of the pairs, the respondent must indicate whether the products are the same or different. Upon completion of the data collection, you compute estimates of δ (called *d′*), as well as τ, for each of the six possible product comparisons. The results are shown in Table 2 (values estimated using the *IFPrograms*™ software). You notice that while the *d′* values are quite different across the product pairs (ranging from 0.37 to 1.53), the estimates of τ are remarkably constant. This illustrates the fact that the size of the difference above which consumers will consider samples as "different" is an internal construct independent of the actual products being evaluated and agrees with previously published results[9]. Using the data in Table 2, you decide to set your criterion at 0.80 units of δ, which are perceptual standard deviation units.

Pair	*d′*	τ
A vs. B	0.56	0.82
A vs. C	1.18	0.79
A vs. D	1.53	0.82
B vs. C	0.70	0.80
B vs. D	0.96	0.79
C vs. D	0.37	0.81

Table 2. Estimated *d′* and τ values from the same-different investigation.

Consumer Relevance Criterion: Your experimentation allowed you to establish the size of the difference relevant to your consumers and you plan on using it in future investigations. A δ value of 0.80 is thus the consumer relevance criterion you would not want an actual product difference to exceed in order to accept an ingredient change.

Your next step is to see how you can use this information in your daily investigations. If you were routinely using consumers, you could simply use the criterion of 0.80. However, you largely use your internal panel to qualify new ingredients and suppliers. Since training can increase sensitivity to product differences through the reduction of perceptual noise[10], you need to translate the consumer criterion into your trained panel scale. You have previously established that your internal panelists are more sensitive than your consumers by approximately 40%, which allows you to translate the criterion of 0.80 into a δ of 1.12 (0.80*1.4). Based on this, you confirm that your current effective panel size of 50 (25 panelists replicated) provides you sufficient power (80%) to detect sensory differences corresponding to δ values equal to 1.12 or greater when using the tetrad test.

Conclusion: Thurstonian modeling of data from the same-different method allowed not only the estimation of the size of the sensory difference, δ, but also a measure of its consumer relevance, τ. This information provides a unique insight into the size of the difference that is relevant for a group of subjects, such as the users of a particular product category. Once τ has been estimated, this value can be used as a reference point against which future results can be assessed. Sensory discrimination programs can also be developed with sample size considerations based on the reference point provided by the estimate of τ.

References and Notes

1. The probability that a difference on a continuous scale assumes a particular value is zero.
2. Chew, V. (1977). Statistical hypothesis testing: An academic exercise in futility. Proceedings of *Florida State Horticultural Society*, **90,** 214–215.
3. Ennis, D. (1990). Relative power of difference testing methods in sensory evaluation. *Food Technology,* **44**(4), 114–117.
4. Ishii, R., Kawaguchi, H., O'Mahony, M., and Rousseau, B. (2007). Relating consumer and trained panels' discriminative sensitivities using vanilla flavored ice cream as a medium. *Food Quality and Preference,* **18**(1), 89–96.
5. Rousseau, B. (2010). Action standards in a successful sensory discrimination program. *IFPress,* **13**(4), 2–3.
6. Macmillan, N., Kaplan, H., and Creelman, C. (1977). The psychophysics of categorical perception. *Psychological Review,* **84**(5), 452–471.
7. Ennis, D., and Ashby, F. (1993). The relative sensitivities of same-different and identification judgment models to perceptual dependence. *Psychometrika,* **58**(2), 257–279.
8. O'Mahony, M., and Rousseau, B. (2002). Discrimination testing: A few ideas, old and new. *Food Quality and Preference,* **14**(2), 157–164.
9. Rousseau, B., Rogeaux, M., and O'Mahony, M. (1999). Mustard discrimination by same-different and triangle tests: Aspects of irritation, memory and τ criteria. *Food Quality and Preference,* **10**(3), 173–184.
10. Rousseau, B. (2004). Relating expert and consumer sensitivities. *IFPress,* **7**(2), 2–3.