Reprinted from IFPress (2001) 4(2)2,3

How Retasting Can Improve the Power of Product Testing

Benoît Rousseau and Daniel M. Ennis

Background: The cost and timing of product testing depends on the size of the experiment. The power of a difference test depends on the number of judgments, a component of the experiment's size. Power is the probability of correctly concluding that a difference exists given a significance level, sample size and specified difference. Efforts to improve the power of difference tests are important because they are rewarded with the development of more resource-efficient methods. Theoretical and applied research on difference testing has shown that some methods may require as much as 100 times the sampling of other methods to detect the same difference with the same power[1]. Some recent experimental work in taste research[2] has shown that retasting during an experiment may improve test power. In order to exploit this finding, it is useful to understand why this might occur. In this report, we develop a possible explanation of this effect and from this model we can predict a broad range of experimental outcomes.

Scenario: One of your main responsibilities is to investigate the sensory characteristics of yogurt products following a process or formulation change. You use the 2-AFC (2-Alternative Forced Choice) method routinely in discrimination testing to evaluate such product modifications. A sensory attribute, such as smoothness, is first chosen prior to the test. Sets of the two alternatives are presented to each member of your panel who is instructed to select the more intense sample on this attribute.

You are interested in investigating the effect of retasting samples in the 2-AFC task. On the one hand, performance might improve by providing more information on the product attribute before generating a response. On the other hand, performance might decrease due to sensory fatigue and confusion. You conduct an experiment in which panelists are tested in two conditions. In the first condition, no retasting is allowed. In the second condition, the panelists are required to retaste each sample once before responding. Table 1 gives a summary of the results.

Table 1. 2-AFC results with and without retasting.

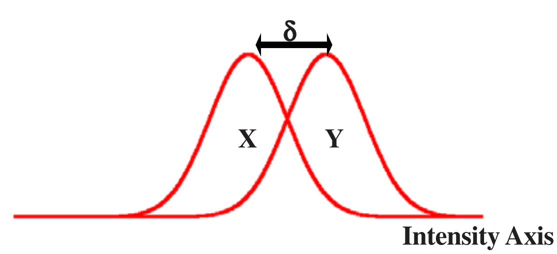
	No. of Correct Responses	Total No. of Responses	d′	Variance of d′
No retasting	138	200	0.7	0.017
One retasting	156	200	1.09	0.02

d´ values are significantly different at $p < 0.05$

You observe that retasting appears to increase the discrimination ability of your panelists. These results confirm published work on retasting[2].

Variance in Models for Difference Tests: Many Thurstonian models for discrimination testing assume an equal variance normal distribution of percepts for each product. Models using this simple assumption have proven to be very useful in explaining many experimental findings, such as Gridgeman's paradox[3]. The difference between the products is called δ and its experimental estimate is d´. The units of δ are perceptual standard deviations (Figure 1.) If the perceptual variance of a product increases, judgments made about that product will be more uncertain and since the units of δ are perceptual standard deviations, the size of δ and the precision of the experiment will decrease. Alternatively, if the perceptual variance decreases, the size of δ will increase, thus improving the statistical power of the experiment. There are several factors that may affect perceptual variance. Some of these factors might induce its decrease (familiarity with the experimental protocol, degree of training provided to a subject, inherent sensitivity of individuals, and extent of retasting allowed[2]), while some of these factors might trigger its increase (influence of sensory adaptation and irritation[4,5,6], extent to which the memory trace for a product changes during a test[7], and number of memory traces interfering in memory[7].) As we just mentioned, the effect of retasting can be seen as reducing the variance of the perceptual distributions used to make a product testing decision. Table 1 gives d′ values for the two experiments previously described and also shows that these values are significantly different from each other; retasting increases the size of d′. The d′ values, their variances and the test conducted on them were obtained using the *ART!* software.

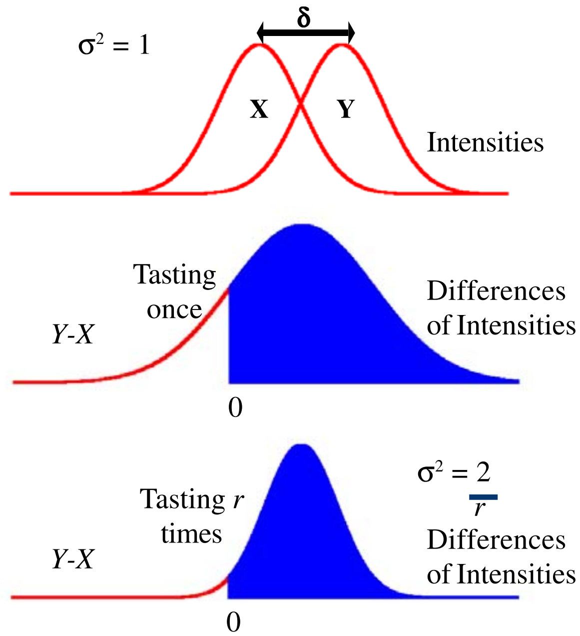
Figure 1. Probabilistic representation of the degree of difference δ between two products X and Y.



Effect of Retasting on Perceptual Variance: Figure 2 illustrates how the variance might decrease as a result of retasting. Consider two products X and Y with means μ_x and μ_y and variance of 1. Since the variance is 1, $\delta = \mu_y - \mu_x$. When sampling each product once, the panelist will experience one sensation for each product, *x* and *y*. In order to make a decision, *y* will be compared to *x* and the product perceived as greatest in intensity will be selected. This decision rule can be described as selecting *y* when *y-x* is positive, and *x* when *y-x* is negative. The probability of this choice can be seen clearly in terms of the distribution of differences between the sensations as shown in Figure 2. This distribution has a variance of 2 (sum of the individual distribution variances.) The subject will be correct when *y-x* is positive; the proportion of correct answers is the blue area under the curve.

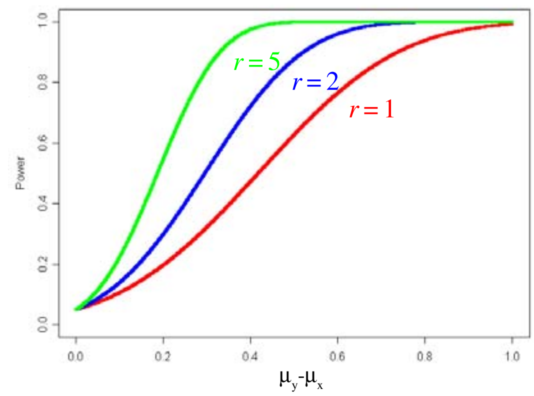
As a panelist retastes the samples, the variance of the difference distribution will decrease[8]. If the samples are tasted *r* times, the variance will decrease by a factor of *r* if the panelist uses an average value of the sensations from the *r* tastings. From the third distribution in Figure 2, we see that the proportion of times *y-x* is positive increases, thus yielding a higher proportion of correct answers. In the absence of other factors that may interfere with performance, this model predicts that performance will improve upon retasting and that the value of δ with *r* tastings will equal $\sqrt{r}$ times δ with no retasting. The extent of improvement depends on how much information can be remembered and used in the decision-making process.

Figure 2. Perceptual variance: Effect of resampling in the 2-AFC method.



Effect of Retasting on Power: Figure 3 shows the power of the 2-AFC method using a sample size of 50 as a function of the sensory difference, $\mu_y - \mu_x$, and the number of tastings, *r*. When *r* is 1, $\delta = \mu_y - \mu_x$. As *r* increases, the perceptual variance decreases leading to greater power at a given value of $\mu_y - \mu_x$. For instance, when $\mu_y - \mu_x$ is 0.5, Figure 3 shows that the probability of declaring a significant difference at $\alpha = 0.05$ is almost 1 when the samples are tasted five times. Contrast this with the power of about 62% when the samples are tasted once. Although these results are specific to the 2-AFC, improvements due to retasting in the power of other methods, such as the triangular and duo-trio methods, can also be predicted and tested.

Figure 3. Power as a function of sensory difference and number of tastings (*r*).



Conclusion: Retasting may increase the power of a difference testing experiment by decreasing the size of the variance of the information used to make a decision. Since sensory differences are expressed in standard units (δ) which are the perceptual standard deviations, a consequence of retasting may be to increase the size of δ. This effect is very general since it applies to many sensory evaluation methods[2] and suggests that retasting will improve the discrimination power of a sensory protocol provided that sensory fatigue does not reduce performance.

When conducting methodological research in order to optimize the reliability of experimental outcomes, it is essential to consider any variable that can hinder or improve the measurement of products' sensory characteristics. Comparing the performance of various protocols using models of the discrimination process will provide valuable information concerning those variables. Probabilistic or Thurstonian models identify and quantify these effects and provide a basis for developing a general foundation to interpret product testing measurements.

References:

1. Ennis, D.M. (1993). The Power of Sensory Discrimination Methods. *Journal of Sensory Studies*, **8**, 353-370.
2. Rousseau, B., & O'Mahony, M. (2000). Investigation of the effect of within-trial retasting and comparison of the dual-pair, same-different and triangle paradigms. *Food Quality and Preference*, **11**, 457-464.
3. Ennis, D.M. (1998). Thurstonian scaling for difference tests. *IFPress*, **1(3)**, 2-3.
4. Rousseau, B. & O'Mahony, M. (1997). Sensory difference tests: Thurstonian and SSA predictions for vanilla flavored yogurts. *Journal of Sensory Studies,* **12**, 127-146.
5. Rousseau, B., Meyer, A. & O'Mahony, M. (1998). Power and sensitivity of the same-different test: Comparison with triangle and duo-trio methods. *Journal of Sensory Studies,* **13**, 149-173.
6. Rousseau, B., Rogeaux, M., & O'Mahony, M. (1999). Mustard discrimination by same-different and triangle tests: Aspects of irritation, memory and t criteria. *Food Quality and Preference*, **10**, 173-184.
7. Rousseau, B., Stroh, S. and O'Mahony, M. (2001). Investigating more powerful discrimination tests with consumers: Effects of memory and response bias. *Food Quality and Preference*, *In Press.*
8. Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswikian origins of uncertainty in judgment: A sampling model of confidence in sensory discrimination. *Psychological Review*, **104**, 344-366.