Food Sci. Biotechnol. Vol. 13, No. 6, pp. 841 ~ 847 (2004)

MINIREVIEW

Food Science and Biotechnology

Sensory Difference Testing: Thurstonian Models

Hye-Seong Lee and Michael OMahony*

Department of Food Science & Technology, University of California, Davis, CA, 95616 USA

Abstract This brief review introduces the concepts and ideas involved in Thurstonian modeling as applied to sensory difference measurement. It summarizes the relevant literature concerned with the theorizing and confirmation of the model. It introduces the concept of stimulus variability and the fundamental measure of sensory difference, d'. It indicates how the paradox of discriminatory non-discriminators can be simply explained using the model. It considers how memory effects and the complex interactions in the mouth can reduce d' by increasing the variance of sensory distributions.

Keywords: sensory evaluation, difference tests, Thurstonian models, Signal Detection Theory, cognitive strategies

Introduction

For a science to progress, it is important to have theories and models. Until recently, the area of sensory and consumer testing was merely a collection of methods without any theoretical backbone. It was optimistically hoped that if the statistical analysis was sophisticated enough, the problems of measurement would vanish. However, more recently, the error of this approach has been realized. Researchers in this area are beginning to realize that the main problems are associated with the act of measurement itself. Therefore, they have begun to pay attention to models of measurement. The current most comprehensive system is Thurstonian modeling.

Sensory difference testing Thurstonian modeling was first applied to the area of difference testing. Sensory difference tests are measurement methods designed for determining the degree of difference between confusable stimuli. Such stimuli are so similar that it is difficult to determine whether they are the same or different. These measurements are important for quality control, determining the effects of ingredient change, processing change or changes in packaging. They are used in storage studies and product development studies involving product imitation.

As a first step, research has shown that not all difference test methods have the same sensitivity. Differences may be found using one test method when they are not found using another. There have been many studies comparing judges' performance on several difference test protocols (1-16). It is because of such studies that the necessity for theories and models became apparent.

As a first basis for theory, Thurstonian modeling (17, 18) and its close cousin, Signal Detection Theory (19, 20), have been applied. Thurstonian models were first applied to the 2-AFC, triangle, and duo-trio discrimination methods (21). Univariate and multivariate models were further developed for a variety of discrimination tests (22-47). Such models have been used to produce tables of d', the Thurstonian index of degree of difference, which can be computed from the proportions of correct responses for the various sensory difference tests (21, 34, 48-52). The application of Thurstonian modeling has been reviewed (24, 53-56). To understand these models, it is important to understand the concept of stimulus variability and the index of discriminability: d'.

Stimulus variability and d' Consider a person tasting an unsweetened cookie. The nerves will transmit all the stimulation elicited by the cookie to the brain. Yet, this stimulation will not be constant. The nervous system will fluctuate in the number of nerves firing messages to the brain and their intensity of firing. There will be complex interactions in the mouth. The taste ingredients of the cookie will be diluted by the fluids in the mouth. These will comprise of saliva, secreted at varying rates, as well as residual stimulus not properly expelled after prior tastings. The taste system will desensitize to these ingredients, which will alter the sensitivity of the taste system and the stimulation transmitted to the brain. In addition, the food product itself may be heterogeneous, which would also add variability to the stimulation elicited by the cookie. Thus there will be variability in the stimulation transmitted by the nerves to the brain that is elicited by the cookie. Sometimes the stimulation will be high; sometimes it will be low. Yet there will be an average level that will occur mostly. The frequencies of occurrence of various levels of stimulation can be represented by a frequency distribution, like the one shown in Fig. 1.

So the first step in Thurstonian modeling is that a food sample can be represented by a distribution of neural input.

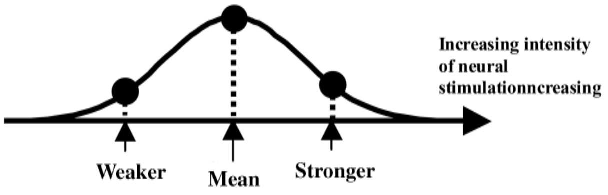

Fig. 1. Figure indicating the frequency at which different levels of neural stimulation are transmitted to the brain for a given stimulus such as an unsweet cookie.

---

*Corresponding author: Tel: 1-530-756-5493; Fax: 1-530-756-7320
E-mail: maomahony@ucdavis.edu
Received September 15, 2004; accepted October 12, 2004

Now consider an easily perceptible amount of sweetener added to the cookie. The sweetener will noticeably increase the amount of stimulation transmitted to the brain, which will be added to the variable input elicited by the rest of the cookie. In terms of our frequency distributions, it can be seen as creating a new distribution further up the intensity axis, by an amount equal to the increased sweetness. This is illustrated at the top of Fig. 2(a). In the figure, the unsweet cookie is generating less neural input, while the sweet cookie is generating a noticeably greater intensity of stimulation (distribution further up the intensity axis, to the right). A person can easily distinguish between the two cookies; the two distributions are far apart.

Yet, if the amount of sweetener added had been so small that the unsweet and sweetened cookies were almost indistinguishable, then the two distributions would be closer to each other. In fact they would overlap; the stimuli would be confusable. This is illustrated by the middle part of Fig. 2(b). At high intensities of sweetness, the cookie will be perceived as sweetened. At low intensities, it will be perceived as unsweetened. At medium intensities, where the distributions overlap, the cookies may be perceived as sweetened or unsweetened. They will be confusable.

Obviously, the further apart the two distributions, the easier it would be to discriminate between the two cookies. Thus, the degree of separation could be used as a measure of how well people could discriminate between the cookies. The distance between the two distributions is called d'; it is a symbol adopted from communications engineering. The larger the value of d', the more discriminable the two stimuli. The distance is measured in terms of standard deviations. Thus, a d' value of 1.5 would indicate that the stimulus distributions were 1.5 standard deviations apart. Basically, d' is a measure of how different two stimuli are, in terms of the variability one stimulus. An engineer would call it a signal-to-noise ratio. It is difficult to get a feel for d' without using it. Yet, a d' of unity represents the difference between two stimuli that can just begin to be distinguished. Their difference is at the threshold level. This can be best understood in terms of a paired comparison difference test. Imagine a judge being asked which of the two cookies appeared to taste sweeter. If the judge could distinguish easily between the stimuli, he would get all his paired comparison tests (100%) correct. If he could not distinguish the two foods, he would guess and get half the tests (50%) correct just by guessing. If his performance was 75%, he would be on the threshold of perceiving a difference; he would be half way between guessing and perfect discrimination. A d' of unity is equivalent to getting 76% of tests correct. To understand further, a d' of 1.5 is equivalent to 86% tests correct, a d' of 2 equivalent to 92%, d' of 2.5 equivalent to 96%, and a d' of 3, 98%.

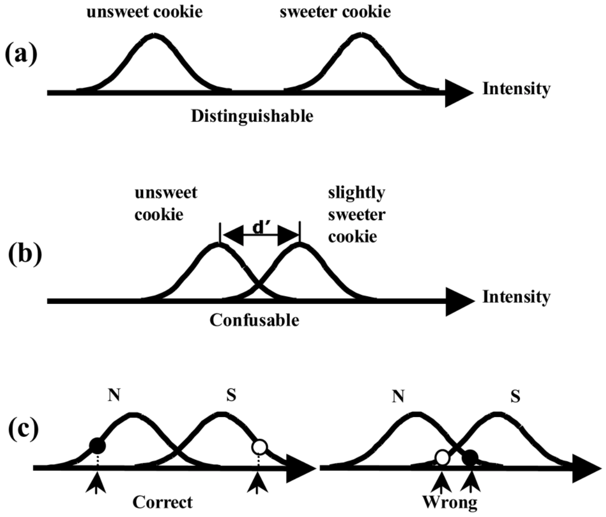

Fig. 2. Food stimuli represented by Thurtonian frequency distributions.

Now consider the bottom part of Fig. 2(c) bottom left. Here are represented the overlapping distributions from two confusable foods 'N' and 'S'. They might be the unsweet and sweet cookies or, to generalize, they might be any two foods or products that are confusable. Let N be the less intense of the two and 'S' be slightly more intense, because of a greater amount of some ingredients or a greater strength of some attribute or attributes. Consider a judge tasting the food 'N'. The intensity at the instant of tasting could be at any level within the range of the 'N' distribution. The tendency will be to have an intensity at the center of the distribution, where the distribution (frequency of occurrence) is at its peak. Yet, it could be at any point within the range of the distribution. Imagine it is at an intensity represented by the black dot. Now, imagine the judge tastes the slightly more intense food, 'S'. This time the momentary intensity is at a level represented by the white dot. If the judge were asked which cookie was sweeter, he would reply that it was the one represented by the 'S' distribution (white dot) and he would be scored as correct.

Now, consider the diagram in section (c) to the right. Here, the nervous system was rather active when food 'N' was tasted, so the level of stimulation was high on the 'N' distribution for that cookie (see black dot). Now, imagine that the nervous system was not so active when the slightly more intense food was tasted. The level of stimulation was towards the low end of the 'S' distribution (white dot). If the judge was asked now which food was stronger tasting, he would choose the food represented by the 'N' distribution, because it was further up the intensity axis and he would be scored as wrong.

So, sometimes the judge is right, and sometimes he is wrong. This is exactly what happens with confusable stimuli, the type of stimuli that are considered in a difference test. The closer the two distributions (smaller d'), the more they will overlap and the greater the likelihood that the judge will give the wrong response. If this exercise were repeated 100,000 times, using Monte Carlo modeling on a computer, it would give a good estimate of the proportion of times that the test would be performed correctly for that particular d'. This could then be repeated for a whole selection of d's resulting in a table of proportions of tests performed correctly for various d' values. More importantly, it would result in a table of d' values for a given proportion of tests correct (49).

This exercise can be repeated for various difference tests, producing tables of d'. This means that a test like the paired comparison and the triangle test can be compared directly, even though their chance probabilities are different. This allows comparisons between tests with different chance probabilities, which were not possible before. d' is a fundamental measure, which is independent of the method used to measure it.

The paradox of discriminatory non-discriminators. The results become interesting for the triangle and 3-AFC methods. For the triangle test, three stimuli are given to the judge, two of which are the same, while the third is different. The judge is required to taste all three and indicate which one is the odd or different one. The 3-AFC method is slightly different. The same three stimuli may be given to the judge, but this time he is told the nature of the difference. For example, he may be told that two of the foods were not sweet while the third one was sweet. After tasting the three stimuli, the judge indicates which one was the sweet stimulus. It turns out that performance on the 3-AFC is superior to that with the triangle method. This was first noted in 1937 by Abrahams and his coworkers (57) and remained unexplained for many years. It was later named the paradox of discriminatory non-discriminators by Gridgeman (8). It might be thought that performance on a 3-AFC would be superior, because the judge knew what attribute he should be concerned with. However, this is not the explanation. With other test methods, judges can perform better when they are not given any indication about the attribute that is changing (58). It was not until 1979 that Frijters (35, 36) explained the paradox, using the concepts of Thurstonian modeling. His explanation was as follows.

Consider the distributions of two foods ‘N’ and ‘S’, shown in Fig. 3. Let there be two less intense ‘N’ stimuli and one slightly more intense ‘S’ stimulus. Thus, two stimuli, indicated by black dots, will be associated with the N distribution and one stimulus (white dot) with the ‘S’ distribution. Consider the top part of the Fig. 3(a). The intensities of the two ‘N’ stimuli (black dots) can be seen to be close to the middle of their distribution. The greater intensity of ‘S’ is indicated by the white dot. If the judge were asked the triangle question, to indicate the odd sample, he would choose the ‘S’ (white dot), because it was further away from the two stimuli represented by the black dots, which themselves were close together. If he were asked the 3-AFC question, to identify the strong stimulus, he would chose the ‘S’ stimulus (white dot), because it was further up the intensity axis. Therefore, both the triangle and the 3-AFC would be scored as correct.

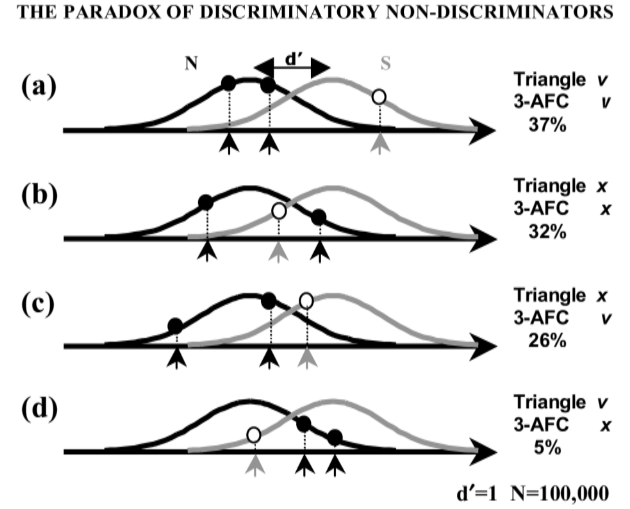

Fig. 3. A Thurstonian analysis for the triangle and 3-AFC tests.

Now, consider the distributions given in row (b). This time, one of the stimuli associated with the ‘N’ distribution is close to the low end of the distribution, while the other is close to the high end. The stimulus associated with the ‘S’ distribution is at the low end of its distribution. In fact, it has a lower intensity than the black dot to its right. In this case, if the judge were performing a triangle test and indicating the ‘odd’ stimulus, he would pick the stimulus associated with the left hand black dot, because it was a greater distance away from the other two. With the 3-AFC method, identifying the most intense stimulus, the stimulus furthest up the intensity axis would be chosen. Accordingly, the judge would pick the stimulus associated with the right hand black dot. In both cases, the triangle and 3-AFC test would have been scored as wrong. Yet, the mistake that was made was not the same in each case. It would appear that the two methods are not equivalent.

Now consider row (c). In this case, one of the ‘black dot’ stimuli is at the low end of its distribution, while the other is towards the high end. The ‘white dot’ stimulus is slightly towards the low end of its distribution. When asked the triangle question, the judge would indicate the left hand ‘black dot’ stimulus as the ‘odd’ one, because it is furthest away from the other two. Accordingly, the triangle test would be scored as incorrect. If asked the 3-AFC question, the judge would select the ‘white dot’ stimulus because it is furthest up the intensity axis. This test would be scored as correct. Therefore, it can be seen that for exactly the same information reaching the brain, the judge will perform the 3-AFC test correctly and the triangle test incorrectly. Yet, the momentary sensitivities of the judge for the two tests are exactly the same. The discrepancy is due to the fact that the judge was making his decision in a different way. The decision-making for the 3-AFC test was more efficient.

Consider the situation shown in row (d). In this case, the stimuli have swapped over completely. This would be a rare occurrence. Yet, in this case, it can be seen that the triangle test would be correct, while the 3-AFC would be scored as wrong.

When the situation was modeled, for a d’ of unity and a sample size of 100,000, the situation illustrated in case (a) occurred 37% of the time. In case (b), the situation occurred 32% of the time. For (c), it was 26%, and for the rarely occurring case (d), it was 5%. Thus, overall, the performance on the 3-AFC test will be superior to that on the triangle test for the same perceived degree of difference (same d'). Frijters’s (35) explanation for the paradox required experimental confirmation. Experiments needed to demonstrate that the same judges, tasting the same stimuli, would perform a higher proportion of 3-AFC tests correctly than triangles, yet would have the same d' in both conditions. This was confirmed by several studies (59-63).

For example, for a d' value of unity, a judge would be expected to perform 42% of triangle tests correctly but 63% of 3-AFCs; for a d' value of two, the proportions of tests correct would be 60 and 87%, respectively. These values are available from tables (49). Yet, the chance level for the both test is 33.33%. Statistical tests such as chi square or the binomial test, based on such chance levels, cannot deal with the fact that performance on the 3-AFC is superior. An analysis based on Thurstonian modeling and d' would be more revealing. Tests of significance of d' are also available (64).

The paired comparison and duo-trio methods can be modeled in the same way. Here, performance on the paired comparison is superior to that on the duo-trio, even though the d' values are the same. Thus, for a d' value of unity, a judge would be expected to perform 58% of the duo-trio tests correctly, while performing 76% of correct paired comparisons; for a d' value of two, the proportions of tests correct would be 75 and 92%, respectively. These values are available from tables (49). Yet, the chance level for both tests is 50%.

Cognitive strategies Why was performance superior for the paired comparison and 3-AFC methods? The reason can be surmised from the decision strategies used for the triangle and 3-AFC methods, illustrated in Fig. 3. For the triangle test, the judge was seeking the stimulus that was furthest away from the other two. He was comparing the distances between the stimuli on the intensity axis. This cognitive strategy, called the 'comparison of distances' strategy, is illustrated in Fig. 4. The same strategy is used for the duo-trio method. For the 3-AFC, the judge chose the most intense stimulus. This is called the 'skimming' strategy, because the most intense stimulus is skimmed off in the same way that froth is skimmed off the top of the beer (Fig. 4). The paired comparison method uses the same cognitive strategy. As could be seen from Fig. 3, the skimming strategy was more efficient than the comparison of distances strategy.

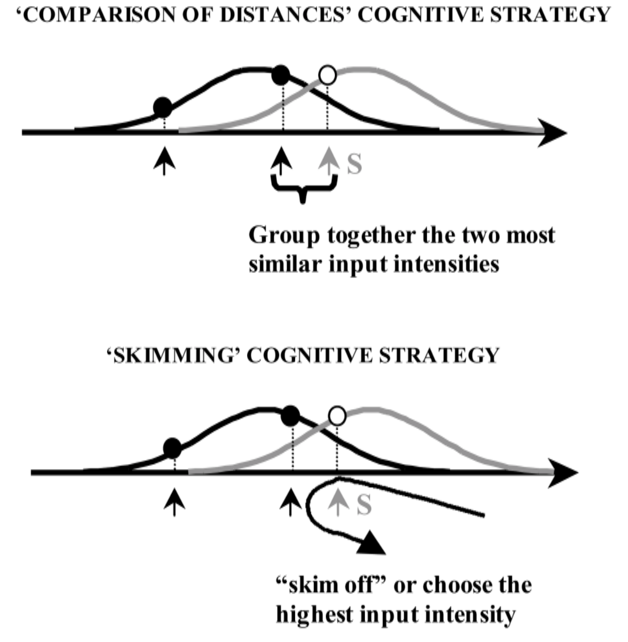

Fig. 4. Illustration of the 'comparison of distances' strategy used in the triangle and duo-trio methods and the 'skimming' strategy used in the paired comparison and 3-AFC methods.

A lack of understanding of these principles can lead to wrong conclusions. If a taste panel performs 90% of their paired comparisons correctly, it might be concluded that the difference between the products being tested was greater than when the panel performed 80% of their duo-trios correctly. This would seem reasonable from a statistical point of view, because the chance probabilities of both tests are the same. Yet, the paired comparison uses the more efficient skimming strategy, while the duo-trio uses the less efficient comparison of distances strategy. To perform 80% of duo-trio's correctly requires a d' of 2.36. To perform 90% of paired comparisons correctly only requires a d' of 1.81. So, it can be seen that, despite the expectation derived from probabilities of guessing (null hypothesis), the difference measured by the duo-trio method was greater than that measured by the paired comparison method.

Statistical power The lack of efficiency of the comparison of distances cognitive strategy can be illustrated by considering the statistical power of these tests. The results of modeling indicate that for a d' value of unity, at the 5% level of significance, using a power of 90% (requiring differences between the products to be detected 90% of the time), the 3-AFC method would require 21 tests to reach significance, while the triangle method would require 276 tests. The comparative figures for the paired comparison and duo-trio methods are 27 and 310, respectively. These values are available from tables (49). This illustrates clearly the massive differences in statistical power between tests using the skimming strategy and the comparison of distances strategy. The traditionally chosen triangle and duo-trio methods are not powerful. They need very large samples of data to detect small differences. They are therefore expensive to perform.

Distribution variance Thurstonian modeling provides further insights by comparing the variances of the sensory distributions for the two food products being tested. Consider once again, the two distributions for the products 'N' and 'S' (Fig. 5). The unit of measurement of d' is the standard deviation. If the standard deviation is increased, then for a given sensory distance, the value of d' will decrease. Thus, for a standard deviation of unity, d' might have a value of two. If the standard deviation were to be doubled, the unit of measurement would become twice as big, and the d' value halved (d'=1). Thus any experimental variable that increases the standard deviation or variance of the distributions will decrease d'. What are these variables?

The reactivity of the taste system can be thought of as creating the perceptual distance between the two distributions. A judge, who has more variability in his nervous system because of large fluctuations in the firing rate of nerves that are not actually involved in indicating the presence of the stimulus under consideration, but are just firing spontaneously, will have a larger variance for his distributions. Consequently, he will have a smaller d'. Another way of expressing this is to say that for two

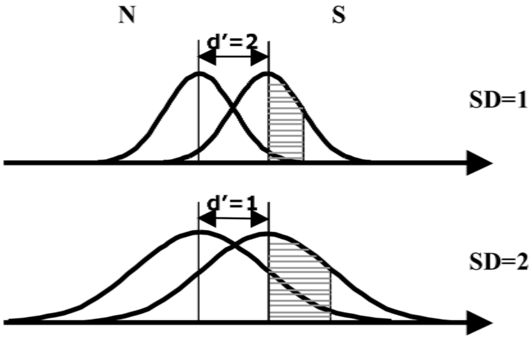

Fig. 5. Sensory distributions indicating the effects of variance change on d'.

products to be perceived as different, they must be more different than any random fluctuations in the sensory system for a single product.

The complex interactions taking place in the mouth have already been discussed and can provide a large source of variance. The greater this variance, the more difficult it will be for the judges to detect differences between products. Similarly, if the products, 'N' and 'S', themselves vary a great deal, because of bad quality control, it will be more difficult for judges to determine differences between the 'S' and 'N' products. For detection, differences between products must exceed the differences within products.

Another source of variance is a result of the sequence of tasting the stimuli. For example, if a sample of food is tasted immediately after a second sample of a food with a similar or stronger intensity, the desensitization caused by some of the first food remaining in the mouth after swallowing will cause the second stimulus to taste weaker than it should (if 'N' is tasted after 'S', it may taste weaker than it should). The effect will be reversed when the order is reversed. Yet, an increase in salivation, or spitting or swallowing ability, could dilute the first stimulus and decrease this effect. Either way, the effect will fluctuate, adding variance to the distributions. The consequences of these effects have been demonstrated and reviewed (31, 60-63, 65-73).

Forgetting and memory effects also provide a source of variance. A difference test involves the comparison of the taste of a food with the memory of a previously tasted food. The greater the time interval between the tastings, the less accurate the memory of the prior food will be (74, 75). Another way of describing this is to say that the greater the time interval between tastings, the greater will be the variance associated with the memory of the previously tasted food. This greater variance will elicit lower d' values, rendering three-stimulus protocols less discriminatory than two-stimulus protocols (13, 71, 72, 76).

Overdispersion The sources of variance discussed above concern perceptual variance, which can reduce d' values. It is also worth briefly considering a different type of variance: the variance of d' itself when statistical analyses are involved. Significance testing for discrimination methods generally involves binomial statistics. Such analyses assume that the discrimination ability of judges is distributed in such a way that it would not distort the 'chance' binomial distribution of performance scores. However, the judges' sensitivities are often distributed in such a way that distortion does occur. The extra variance added by the distortion in the distribution of the judges sensitivities, away from a binomial distribution, is called overdispersion and must be accounted for. Overdispersion increases the variance of d' values, which can lead to Type I errors if unaccounted for. The extra variance makes it necessary to have larger samples of data to demonstrate significant differences between stimuli. Because of overdispersion, a beta-binomial analysis (77-81), which compensates for overdispersion, is more suitable than a simple binomial analysis for testing the significance of forced choice tests. However, a beta-binomial analysis requires that each judge should perform replicate tests for each method. It is also worth noting that the variance of d' also tends to vary with its size. This is described by psychophysical power functions (24).

Conclusion From this brief introduction, it can be seen that Thurstonian modeling gives insight into several aspects of difference testing. Perhaps the most important point is the idea of stimulus variability. A food that is tasted repeatedly, will not taste identical on each tasting. Stimulus variability is the first fundamental of Thurstonian modeling. A consideration of this variability provides a fundamental measure of the difference between two confusable stimuli, d', which is independent of the method used to measure it. Such fundamental measures are essential for any science. Measures of mere performance, like the proportion of tests correct, are not independent of the method. The performance on a 3-AFC test will be superior to that for a triangle test, while the sensitivity (d') for both these tests can be seen to be the same. Finally, a consideration of the variance of distributions provides an explanation for experimental variables associated with memory or the sequence of tasting.

The examples given above are univariate and concern difference tests. The same ideas have been extended to intensity scaling, multivariate difference tests, multivariate statistics, and product mapping. Thurstonian models provide a powerful structure for understanding the mechanisms of sensory measurement.

References

1. Braun V, Rogeaux M, Schneid N, OMahony M, Rousseau B. Corroborating the 2-AFC and 2-AC Thurstonian models using both a model system and sparkling water. Food Qual. Pref. 15: 501-507 (2004)
2. Byer AJ, Abrams D. A comparison of the triangular and two-sample taste-test methods. Food Technol. 7: 185-187 (1953)
3. Dawson EH, Dochterman E. A comparison of sensory methods of measuring differences in food qualities. Food Technol. 5: 79-81 (1951)
4. Filipello F. A critical comparison of the two-sample and triangular binomial designs. Food Res. 21: 235-241 (1956)
5. François P, Sauvageot F. Comparison of the efficiency of pair, duo-trio and triangle tests. J. Sens. Stud. 3: 1-94 (1988)
6. Gridgeman NT. Taste comparisons: Two samples or three? Food Technol. 9: 148-150 (1955)
7. Gridgeman NT. Group size in taste sorting trials. Food Res. 21: 534-539 (1956)
8. Gridgeman NT. A re-examination of the two-stage triangle test

for the perception of sensory differences. J. Food Sci. 35: 87-91 (1970)

9. Grim AC, Goldblith SA. Some observed discrepancies in application of the triangle test to evaluation of irradiated whole egg magma. Food Technol. 19: 146 (1965)
10. Helm E, Trolle B. Selection of a taste panel. Wallerstein Lab. Commun. 9: 181-194 (1946)
11. Hopkins JW. Some observations on sensitivity and repeatability of triad taste difference tests. Biometrics 10: 521-530 (1954)
12. Hopkins JW, Gridgeman NT. Comparative sensitivity of pair and triad flavor intensity difference tests. Biometrics 11: 63-68 (1955)
13. Lau S, O'Mahony M, Rousseau B. Are three-sample tasks less sensitive than two-sample tasks? Memory effects in the testing of taste discrimination. Percept. Psychophys. 66: 464-474 (2004)
14. O'Mahony M, Wong SY, Odbert N. Sensory difference tests: Some rethinking concerning the general rule that more sensitive tests use fewer stimuli. Lebensm.Wiss. Technol. 19: 93-94 (1986)
15. Pókorny J, Marcín A, Davídek J. Comparison of the efficiency of triangle and tetrad tests for discriminatory sensory analysis of food. Die Nahrung 25: 561-564 (1981)
16. Wasserman AE, Talley F. A sample bias in the evaluation of smoked frankfurters by the triangle test. J. Food Sci. 34: 99-100 (1969)
17. Thurstone LL. A law of comparative judgment. Psychol. Rev. 34: 273-286 (1927)
18. Thurstone LL. Psychophysical analysis. Am. J. Psychol. 38: 368-389 (1927)
19. Green DM, Swets JA. Signal DetectionTheory and Psychophysics. John Wiley & Sons, New York, NY, USA. pp. 479 (1966)
20. Macmillan NA, Creelman CD. Detection Theory: A User's Guide. Cambridge University Press, Cambridge, UK. pp. 407 (1991)
21. Ura S. Pair, triangle and duo-trio test. Rep. Stat. Appl. Res., JUSE 7: 107-119 (1960)
22. Ennis DM. Confusable and discriminable stimuli: Comment on Nosofsky (1986) and Shepard (1986). J. Exp. Psychol. Gen. 117: 408-411 (1988)
23. Ennis DM. Multivariate sensory analysis. Food Technol. 42: 118-122 (1988)
24. Ennis DM. Relative power of difference testing methods in sensory evaluation. Food Technol. 44: 114-118 (1990)
25. Ennis DM. Modeling similarity and identification when there are momentary fluctuations in psychological magnitude. pp. 279-298. In: Multidimensional models of perception and cognition. Ashby FG (ed). Lawrence Erlbaum Associates, Hillsdale, NJ, USA (1992)
26. Ennis DM, Mullen K. The effect of dimensionality on results from the triangular method. Chem. Senses 10: 605-608 (1985)
27. Ennis DM, Mullen K. A multivariate model for discrimination methods. J. Math. Psychol. 30: 206-219 (1986)
28. Ennis DM, Mullen K. Theoretical aspects of sensory discrimination. Chem. Senses 11: 513-522 (1986)
29. Ennis DM, Mullen K. A general probabilistic model for triad discrimination, preferential choice, and two-alternative identification. pp. 115-122. In: Multidimensional models of perception and cognition. Ashby FG (ed). Lawrence Erlbaum Associates, Hillsdale, NJ, USA (1992)
30. Ennis DM, Mullen K. Probabilistic psychophysics with noisy stimuli. Math. Soc. Sci. 23: 221-234 (1992)
31. Ennis DM, O'Mahony M. Probabilistic models for sequential taste effects in triadic choice. J. Exp. Psychol.Human Percept. Perf. 21: 1088-1097 (1995)
32. Ennis DM, Mullen K, Frijters JER. Variants of the method of triads: Unidimensional Thurstonian models. Brit. J. Math. Stat. Psychol. 41: 25-36 (1988)
33. Ennis DM, Palen JJ, Mullen K. A multidimensional stochastic theory of similarity. J. Math. Psychol. 32: 449-465 (1988)
34. Ennis JM, Ennis DM, Yip D, O'Mahony M. Thurstonian models for variants of the method of tetrads. Brit. J. Math. Stat. Psychol. 51, 205-215 (1998)
35. Frijters JER. The paradox of discriminatory nondiscriminators resolved. Chem. Sens. Flav. 4: 355-358 (1979)
36. Frijters, J.E.R. Variations of the triangular method and the relationship of its unidimensional probabilistic models to three-alternative forced-choice signal detection theory models. Brit. J. Math. Stat. Psychol. 32: 229-241 (1979b)
37. Frijters JER. Three-stimulus procedures in olfactory psychophysics: An experimental comparison of Thurstone-Ura and three-alternative forced-choice models of Signal Detection. theory. Percept. Psychophys. 28: 390-397 (1980)
38. Frijters JER. The excitatory state in the triangular constant method. Psychometrika 46: 219-222 (1981)
39. Frijters, J.E. An olfactory investigation of the compatibility of oddity instructions with the design of a 3-AFC Signal Detection task. Acta Psychol. 49: 1-16 (1981)
40. Kapenga JA, de Doncker E, Mullen K, Ennis DM. The integration of the multivariate normal density function for the triangular method. pp. 321-328. In Numerical Integration: Recent Developments, Software, and Applications. Keast P, Fairweather G (eds). D. Reidel Publ. Co., Dordrecht, Holland (1987)
41. Mosteller F. Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. Psychometrika 16: 3-9 (1951)
42. Mosteller F. Remarks on the method of paired comparisons: II. The effect of an aberrant standard deviation when equal standard deviations and equal correlations are assumed. Psychometrika 16: 203-206 (1951)
43. Mosteller F. Remarks on the method of paired comparisons: III. A test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. Psychometrika 16: 207-218 (1951)
44. Mullen K, Ennis DM. Mathematical formulation of multivariate Euclidean models for discrimination methods. Psychometrika 52: 235-249 (1987)
45. Mullen K, Ennis DM. A simple multivariate probabilistic model for preferential and triadic choices. Psychometrika 56: 69-75 (1991)
46. Mullen K, Ennis DM, de Doncker E, Kapenga JA. Models for the duo-trio and triangular methods. Biometrics 44: 1169-1175 (1988)
47. Vessereau A. Les méthodes statistiques appliquées au test des caractères organoleptiques. Ann. Nutr. 19: 103-140 (1965)
48. Elliott PB. Tables of d'. pp. 651-684. In: Signal Detection and Recognition by Human Observers: Contemporary Readings. Swets JA (ed). John Wiley and Sons, New York, NY, USA (1964)
49. Ennis DM. The power of sensory discrimination methods. J. Sens. Stud. 8: 353-370 (1993)
50. Frijters JE. Expanded tables for conversion of a proportion of correct responses (P_c) to the measure of sensory difference (d') for the triangular method and the 3-alternative forced choice procedure. J. Food Sci. 47: 139-143 (1982)
51. Frijters JER, Kooistra A, Vereijken PFG. Tables of d' for the triangular method and the 3-AFC Signal Detection procedure. Percept. Psychophys. 27: 176-178 (1980)
52. Hacker MJ, Ratcliff R. A revised table of d' for M-alternative forced choice. Percept. Psychophys. 26: 168-170 (1979)
53. O'Mahony M. Understanding discrimination tests: A user-friendly treatment of response bias, rating and ranking R-index tests and their relationship to signal detection. J. Sens. Stud. 7: 1-47 (1992)
54. O'Mahony M. Who told you the triangle test was simple? Food Qual. Pref. 6: 227-238 (1995)
55. O'Mahony M, Masuoka S, Ishii R. A theoretical note on difference tests: Models, paradoxes and cognitive strategies. J. Sens. Stud. 9: 247-272 (1994)
56. O'Mahony M, Rousseau B. Discrimination testing: A few ideas, old and new. Food Qual. Pref. 14: 157-164 (2002)
57. Abrahams H, Krakauer D, Dallenbach KM. Gustatory adaptation to salt. Am. J. Psychol. 49: 462-469 (1937)
58. Yip DH. Triadic and tetradic taste discrimination testing: Thurstonian and sequential effects. MS thesis, University of California, Davis, CA, USA (1996)
59. Delwiche J, O'Mahony M. Flavour discrimination - An extension of Thurstonian paradoxes to the tetrad method. Food Qual. Pref. 7: 1-5 (1996)
60. Masuoka S, Hatjopoulos D, O'Mahony M. Beer bitterness detection: Testing Thurstonian and Sequential Sensitivity Analysis

models for triad and tetrad methods. J. Sens. Stud. 10: 295-306 (1995)

61. Rousseau B, O'Mahony M. Sensory difference tests: Thurstonian and SSA predictions for vanilla flavored yogurts. J. Sens. Stud. 12: 127-146 (1997)
62. Stillman JA. Response selection, sensitivity, and taste-test performance. Percept. Psychophys. 54: 190-194 (1993)
63. Tedja S, Nonaka R, Ennis DM, O'Mahony M. Triadic discrimination testing - Refinement of Thurstonian and Sequential Sensitivity Analysis approaches. Chem. Senses 19: 279-301 (1994)
64. Bi J, Ennis DM, O'Mahony M. How to estimate and use the variance of d' from difference tests. J. Sens. Stud. 12: 87-104 (1997)
65. Dessirier J-M, O'Mahony M. Comparison of d' values for the 2-AFC (paired comparison) and 3-AFC discrimination methods: Thurstonian models, sequential sensitivity analysis and power. Food Qual. Pref. 10: 51-58 (1999)
66. Dessirier J-M, Sieffermann J-M, O'Mahony M. Taste discrimination by the 3-AFC method: testing sensitivity predictions regarding particular tasting sequences based on the sequential sensitivity analysis model. J. Sens. Stud. 14: 271-287 (1999)
67. O'Mahony M. Salt taste adaptation: the psychophysical effects of adapting solutions and residual stimuli from prior tastings on the taste of sodium chloride. Perception 8: 441-476 (1979)
68. O'Mahony M, Goldstein LR. Effectiveness of sensory difference tests: Sequential Sensitivity Analysis for liquid food stimuli. J. Food Sci. 51: 1550-1553 (1986)
69. O'Mahony M, Goldstein LR. Tasting successive salt and water stimuli: the roles of adaptation, variability in physical signal strength, learning, supra- and subadapting signal detectability. Chem. Senses 12: 425-436 (1987)
70. O'Mahony M, Odbert N. A comparison of sensory difference procedures: sequential Sensitivity Analysis and aspects of taste adaptation. J. Food Sci. 50: 1055-1058 (1985)
71. Rousseau B, Meyer A, O'Mahony M. Power and sensitivity of the same-different test: Comparison with triangle and duo-trio methods. J. Sens. Stud. 13: 149-173 (1998)
72. Rousseau B, Rogeaux M, O'Mahony M. Mustard discrimination by same-different and triangle tests: aspects of irritation, memory and τ criteria. Food Qual. Pref. 10: 173-184 (1999)
73. Vié A, O'Mahony M. Triangular difference testing: Refinements to Sequential Sensitivity Analysis for predictions for individual triads. J. Sens. Stud. 4: 87-103 (1989)
74. Avancini de Almeida TC, Cubero E, O'Mahony M. Same-different discrimination tests with interstimulus delays up to one day. J. Sens. Stud. 14: 1-18 (1999)
75. Cubero E, Avancini de Almeida TC, O'Mahony M. Cognitive aspects of difference testing: Memory and interstimulus delay. J. Sens. Stud. 10: 307-324 (1995)
76. Rousseau B, Stroh S, O'Mahony M. Investigating more powerful discrimination tests with consumers: effects of memory and response bias. Food Qual. Pref. 13: 39-45 (2002)
77. Bi J, Ennis DM. A Thurstonian variant of the beta-binomial model for replicated difference tests. J. Sens. Stud. 13: 461-466 (1998)
78. Bi J, Ennis DM. The power of sensory discrimination methods used in replicated difference and preference tests. J. Sens. Stud. 14: 289-302 (1999)
79. Bi J, Ennis DM. Beta-binomial tables for replicated difference and preference tests. J. Sens. Stud. 14: 347-368 (1999)
80. Bi J, Templeton-Janik L, Ennis JM, Ennis DM. Replicated difference and preference tests: how to account for inter-trial variation. Food Qual. Pref. 11: 269-273 (2000)
81. Ennis DM, Bi J. The beta-binomial model: accounting for inter-trial variation in replicated difference and preference tests. J. Sens. Stud. 13: 389-412 (1998)