Food Quality and Preference 10 (1999) 51–58

Food Quality and Preference

Comparison of d′ values for the 2-AFC (paired comparison) and 3-AFC discrimination methods: Thurstonian models, sequential sensitivity analysis and power

Jean-Marc Dessirier, Michael O'Mahony*

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

Received 5 June 1997; received in revised form 21 July 1998; accepted 24 July 1998

Abstract

Using a model system comprising of low concentration NaCl vs water and a beverage system of varying strength apple juice, it was found that the 2-AFC (paired comparison) discrimination method yielded a higher d′ than the 3-AFC method. Although the 3-AFC method is theoretically more powerful than the 2-AFC, the higher d′ of the latter compensates for this, making it the more powerful of the two. © 1999 Elsevier Science Ltd. All rights reserved.

1. Introduction

Sensory discrimination or difference tests are designed for determining the presence or extent of fine differences between very similar, confusable stimuli. There are a range of methods used for discrimination testing; for food discrimination, forced choice procedures are commonly utilised. Comparison of subjects' performance on such tests has been the subject of several studies (Buchanan, Givon, & Goldman, 1987; Byer & Abrams, 1953; Dawson & Dochterman, 1953; Filipello, 1956; François & Sauvageot, 1988; Gridgeman, 1955, 1956, 1970; Grim & Goldblith, 1965; Helm & Trolle, 1946; Hopkins, 1954; Hopkins & Gridgeman, 1955; MacRae & Geelhoed, 1992; Masuoka, Hatjopoulos, & O'Mahony, 1995; O'Mahony, Wong, & Odbert, 1986; Pokorny, Marcín, & Davídek, 1981; Rousseau & O'Mahony, 1997; Stillman, 1993; Tedja, Nonaka, Ennis, & O'Mahony, 1994; Thieme & O'Mahony, 1990; Wasserman & Talley, 1969).

As a step towards the development of some explanatory theory, Thurstonian modelling (Thurstone, 1927a, b) was initially applied to the paired comparison, triangle and duo-trio tests (Ura, 1960). Univariate and multivariate Thurstonian models were further developed for a variety of discrimination tests (Ennis, 1988a,b, 1990, 1992; Ennis & Mullen, 1985, 1986a,b, 1992a,b; Ennis, Mullen, & Frijters, 1988; Ennis, Palen, & Mullen, 1988; Frijters, 1979a,b, 1980, 1981a,b, 1982; Frijters, Kooistra, & Vereijken, 1980; Kapenga, de Doncker, Mullen, & Ennis, 1987; Mosteller, 1951a,b,c; Mullen & Ennis, 1987, 1991; Mullen, Ennis, de Doncker, & Kapenga, 1988; Vessereau, 1965). Such models were used to produce tables of a signal to noise ratio measure of discrimination variously called d′ (Green & Swets, 1966; MacMillan & Creelman, 1991) or δ (Ennis, 1990, 1993). Bi, Ennis, & O'Mahony (1997) denoted d′ as the sample estimate of δ the population value. Tables give d′ or δ values corresponding to proportions of correct responses for the various forced choice procedures (Ennis, 1993; Ennis & Mullen, 1986b; Frijters, 1982; Frijters et al., 1980; Hacker & Ratcliffe, 1979; Ura, 1960). The application of Thurstonian modelling to difference tests has been reviewed (Ennis, 1990; O'Mahony, 1995; O'Mahony, Masuoka, & Ishii, 1994).

Frijters (1979a) used the Thurstonian approach to explain a result observed by Byer & Abrams (1953), later called the paradox of discriminatory nondiscriminators (Gridgeman, 1970). The basis of this paradox was the observation that for a given d′, subjects discriminated correctly a lower proportion of triangle tests (Peryam, 1958; Peryam & Swartz, 1950) than 3-alternative forced choice (3-AFC) tests (Green & Swets, 1966). Yet, they are both triadic protocols which only

*Corresponding author. Tel.: +1-530-752-6389; fax: +1-530-752-4759; e-mail: jadessirier@ucdavis.edu

0950-3293/99/$—see front matter
PII: S0950-3293(98)00037-8

differ slightly in the instructions given to the subject. For the triangle test, the subject is told that two samples are identical, one is different. He is required to find the odd or different sample. In the 3-AFC procedure, the nature of the difference is specified. For example, one product may be sweeter than the other two; the subject must indicate the sweeter sample. Frijters noted that the superior performance found for the 3-AFC was due to a difference in cognitive strategies. During a triangle test, the subject compared the sensory distance between the samples, whereas during a 3-AFC test, the subject was looking for the highest or lowest intensity on a sensory continuum. These two cognitive strategies or decision rules have been called the 'comparison of distances' strategy and the 'skimming' strategy respectively (O'Mahony et al., 1994).

The 2-alternative forced choice (2-AFC) test also uses the skimming strategy and thus for a given d′ value, subjects would perform a higher proportion of tests correctly than a test like the duo-trio, with the same chance level, but using the less efficient comparison of distances strategy (O'Mahony et al., 1994). Ennis (1990, 1993), using Thurstonian modelling, showed that the paired comparison (2-AFC) test and the 3-AFC test are more powerful than the more commonly used duo-trio and triangle tests. By this is meant that for a given d′ value, it needs fewer tests to reject the null hypothesis (H_0: $d'=0$) for the 3-AFC than the triangle method and for the 2-AFC than the duo-trio.

While the Thurstonian approach to explain differences in performance on discrimination tests is well developed, there is a further factor affecting performance. Stimuli are presented in sequences and the sequencing can affect performance. Such effects have been studied and a resulting model developed: Sequential Sensitivity Analysis or SSA. This has been used to predict performance in discrimination tests for a frequently studied model system: 'threshold' NaCl solutions vs purified water (O'Mahony & Goldstein, 1987; O'Mahony & Odbert, 1985; Tedja et al., 1994; Vié & O'Mahony, 1989) as well as for food systems (Masuoka et al., 1995; O'Mahony & Goldstein, 1986; Rousseau & O'Mahony, 1997). The model considers how well each stimulus to be discriminated in a difference test, is detected and identified correctly. Separate experimentation , based on *R*-index measures (Brown, 1974; O'Mahony, 1992; O'Mahony & Odbert, 1985) can be used to measure the relative detectability of the sensations that will be encountered during the difference test.

Consider a difference test being used to test discrimination between two confusable stimuli: 'S' a slightly stronger stimulus and 'W', a slightly weaker stimulus. How well the sensory attributes characterising 'S' are detected upon tasting will depend on what stimulus (or stimuli) was tasted immediately beforehand. The sensory attributes characterising the stronger stimulus 'S' appear easier to detect when 'S' is tasted after the weaker stimulus 'W' (W–S) than when it is tasted after the stronger stimulus (S–S). This can be easily explained as a sensory adaptation effect (O'Mahony, 1979, 1986). The attributes characterising the weaker stimulus 'W' appear to be easier to detect when it is tasted after the stronger stimulus (S–W) than when it is tasted after the weaker stimulus (W–W). These two latter orders are intermediate between the first two. For the 'threshold' NaCl (S)/water (W) model system, the detectability of the characteristic sensory attributes of 'S' or 'W' can be ranked W–S (easiest), S–W, W–W, S–S (most difficult) (O'Mahony & Goldstein, 1987; O'Mahony & Odbert, 1985; Tedja et al., 1994; Vié & O'Mahony, 1989). This ranking is an average value; individuals may vary in their order. Also, the individual rankings for a given subject can change with practice; the system is labile (O'Mahony & Goldstein, 1987). This order of detectability was found to be the result of several interacting factors: sensory adaptation, difference in detectability of supra- and subadapting NaCl stimuli, blending of stimuli with secreted saliva, learning and response bias effects (O'Mahony & Goldstein, 1987). The ranking for the discriminability of various formulations (W vs S) of beverages, beers, wines and yogurts has also been measured and found to be compatible with the ranking obtained with the NaCl/water model system (Masuoka et al., 1995; O'Mahony & Goldstein, 1986; Rousseau & O'Mahony, 1997).

By considering the orders that occur in any difference testing protocol, the relative number of easier and more difficult orders can be used to predict test sensitivity. Hence a triadic test with the stronger stimulus as the 'odd' sample (sequences WWS, WSW, SWW) will be more sensitive (eliciting a higher d′ value) than the one with the weaker stimulus as odd (sequences SSW, SWS, WSS). The latter contains the worst order 'S–S', the former does not. Better performance on the 'S-odd' triad than the 'W-odd' was confirmed for the NaCl/water model system (O'Mahony & Odbert, 1985; Vié & O'Mahony, 1989; Tedja et al., 1994). Better performance on 'S-odd' triads found for foods was also found to be compatible with SSA analysis (Masuoka et al., 1995; O'Mahony & Goldstein, 1986; Rousseau & O'Mahony, 1997). There is also a Thurstonian explanation for this effect based on hypothesising that the stronger stimulus has a larger variance than the weaker stimulus. Current research by the present authors, to be published later, would tend to negate this explanation. A further explanation based on conditional stimulus concepts (Ennis & O'Mahony, 1995) combines both SSA and Thurstonian approaches. This will be discussed later.

The sequencing effect considered by SSA is concerned with changes in sensitivity, d′. The less interference in the strength of the stimuli caused by the conditions in the mouth, the clearer and more discernible will be the sensation evoked by those stimuli. They will, thus, be more discriminable and tests will yield higher d′ values. Thus a triadic test with 'S-odd' sequences yields a higher d′ value than a triadic test with 'W-odd' sequences.

Considering the 2-AFC test, it can also be hypothesised that a 2-AFC (paired comparison) test with the two best sequences (W–S, S–W) should elicit higher d′ values than triadic tests like the triangle or the 3-AFC. This was confirmed by Rousseau and O'Mahony (1997) and will be one of the hypotheses tested in the current study.

Ennis (1990, 1993) using Thurstonian modelling showed how the 3-AFC test is somewhat more powerful than the 2-AFC over various values of d′. Yet, Rousseau and O'Mahony (1997) showed that the sequencing effects encountered in the 2-AFC method elicited a higher d′ than those in the 3-AFC, as predicted by SSA. The higher d′ accordingly increased the statistical power of the 2-AFC. This increase in power more than offset the theoretically predicted advantage of the 3-AFC over the 2-AFC. The 2-AFC was rendered more powerful.

Rousseau and O'Mahony's (1997) study required subjects to discriminate between formulations of vanilla yoghurt. The aim of this study was to generalise this result to other stimuli. Firstly the threshold NaCl/purified water model system was used because there is a sizeable body of knowledge about SSA effects for this system. Secondly, to generalise to other foods, a beverage system was utilised: apple juice.

2. Experiment I

The objective of this experiment was to compare d′ values obtained by 2-AFC and 3-AFC discrimination methods and determine their effect on statistical power. A model system: 'threshold' NaCl vs. purified water was chosen as a medium because SSA effects are well documented for discrimination between these stimuli.

2.1. Materials and methods

2.1.1. Subjects

Twenty subjects (5 M, 15 F; age range 20–36 years), students and staff at U.C. Davis, were tested. Subjects fasted, except for water, for at least 1 h prior to testing. All were naive to the specific aim of the study; 8 had participated in psychophysical experiments before.

2.1.2. Stimuli

Stimuli consisted of 'threshold' NaCl solutions and purified water stimuli. The NaCl solutions were prepared by dissolving reagent grade NaCl (Mallinckrot Inc., Paris, KY) in Milli-Q purified water. The Milli-Q purified water was deionised water fed into a Milli-Q system involving ion exchange and activated charcoal (Millipore Corp., Bedford, MA). The resulting purified water had a specific conductivity of $<10^{-6}$ mho/cm and a surface tension of $\leqslant$71 dynes/cm. The 'threshold' NaCl concentrations varied according to subjects' sensitivity. They ranged 0.3–3.0 mM (0.3 mM, 2 subjects; 0.5 mM, 7 subjects; 0.7 mM, 2 subjects; 1 mM, 4 subjects; 2 mM, 4 subjects; 3 mM, 1 subject).

The samples were dispensed in 10 ml aliquots using Repipet Adjustable Dispensers (Labindustries Inc., Berkeley, CA) in 1 oz. plastic portion cups (S.E. Rykoff & Co., Los Angeles, CA). The whole 10 ml aliquot was sipped and expectorated. All stimuli were presented at constant room temperature (20–22°C), on aluminium trays with the stimuli for six 2-AFC or six 3-AFC on each.

2.1.3. Procedure

Subjects performed twelve 2-AFC and twelve 3-AFC tests in a single experimental session (session length ranged 10–23 min). Preliminary experimentation had indicated that such session lengths were possible without loss of sensitivity. Tests were presented in sets of six (six 2-AFC or six 3-AFC). To compensate for learning or fatigue effects, the presentation of the sets was counterbalanced following a ABBA design. Thus, a subject would perform six 2-AFC tests followed by two sets of six 3-AFC tests followed by six 2-AFC tests (ABBA). Half of the subjects followed the ABBA order, the other half the inverse BAAB order. For the twelve 2-AFC tests, each possible order of presentation (W–S, S–W) was presented six times in random order. For the twelve 3-AFC tests, each possible order of presentation (WWS, WSW, SWW) was presented four times in random order. To prevent the first sample of a test being influenced by the last sample of the prior test, a purified water rinse was taken before each test.

Before each set of six tests, the instructions were repeated. They were for the 2-AFC: "Here are two samples, one is a water and the other is a salt, which one is the salt?" and for the 3-AFC: "Here are three samples, two are water and one is a salt, which one is the salt?". It should also be noted that only 3-AFC tests with the stronger stimulus (NaCl) as the 'odd' sample were performed. This was to elicit a higher d′ value and thus provide a more exacting test for the 2-AFC.

An experimental session started with subjects rinsing 6 times with purified water to cleanse the mouth. This was followed by a warm-up procedure (Pfaffmann, 1954; O'Mahony, Thieme, & Goldstein, 1988) to ensure subjects knew the appropriate signals to be detected and to define the intensity or 'information axis' to be encountered during the experiment. The warm-up consisted of alternate tasting of the two stimuli until the subject felt confident to begin. Initially subjects warmed up with stimuli more discriminable than used in the experiment (NaCl 0.5–2 mM higher than experimental concentrations) and then progressed to using the experimental concentrations. At each concentration, subjects required 3–12 tastings of each stimulus.

Before performing the experimental session, a practice session was conducted to familiarise subjects with the testing procedure and to determine subjects' sensitivity. During this session subjects performed a set of six 2-AFC tests for 3 different concentrations of NaCl (3, 1, and 0.5 mM) with a warm-up prior to each set.

2.1.4. Results

The proportions of correct 2-AFC and 3-AFC tests performed by each subject were used to determine appropriate d′ values (Ennis, 1993). A significant majority of subjects (16/20) had a higher d′ value for the 2-AFC than the 3-AFC method (binomial, $p = 0.02$). A parametric comparison is complicated by the fact that 5 subjects performed all twelve 2-AFC tests correctly, which would give infinite d′ values. Accordingly these subjects were dropped from the calculation leaving fifteen. Mean d′ values for these 15 subjects were 0.9 for the 2-AFC and 0.6 for the 3-AFC. These values were significantly different (*t*-test, $p = 0.04$).

The above *t*-test has the disadvantage of omitting some subjects. As an alternative, the total number of 2-AFC tests performed correctly taken over all 20 subjects could be used to give a 'pooled d′ value' for the whole sample of 20 subjects. Doing this would involve pooling data from replicate tastings and subjects. Certainly this violates the assumptions for a binomial test and introduces overdispersion. However, the problem is solved by using the beta-binomial model (Ennis & Bi, 1998). The same approach for dealing with overdispersion can be used in the calculation of pooled d′ values (Bi & Ennis, 1998). Consequently, pooled d′ values correcting for overdispersion were calculated for both the 2-AFC and 3-AFC tests. They were 1.2 and 0.6 respectively. They were found to be significantly different ($p = 0.005$) using the method of Bi et al. (1997). A further argument for using this approach is that the d′ values obtained for each judge from only 12 tests might have variances large enough to distort the *t*-test result.

Thurstonian modelling indicates that for a given d′ value, a 3-AFC is statistically more powerful than a 2-AFC (Ennis, 1990, 1993). When d′ values vary, test power can be compared by considering the minimum number of tests required to reject H_0 with say, $\alpha < 0.05$ and the probability of rejecting a false H_0 set at say, 0.9. For d′ = 0.6, this would require 75 2-AFC tests and 61 3-AFC tests (Ennis, 1993). Yet for a d′ for the 2-AFC test of 1.2, only 19 tests would be required, indicating how the higher d′ for the 2-AFC more than compensated for the statistical power of the 3-AFC.

3. Experiment II

The objective of this experiment was to compare d′ values for 2-AFC and 3-AFC discrimination tests and determine their effect on statistical power. Apple juice was chosen as a medium to extend the results to a foodstuff.

3.1. Materials and methods

3.1.1. Subjects

Twenty subjects (8 M, 12 F; age range 20–41 years) students and staff at U.C. Davis, were tested. Subjects fasted, except for water, for at least 1 h prior to testing. All were naive to the specific aim of the study; 6 had participated in psychophysical experiments before, although none had participated in Experiment I.

3.1.2. Stimuli

Stimuli were prepared from a single batch of pure apple juice (Tree Top Inc., Selah, Washington, USA). Subjects were required to discriminate between the pure product and juice which had been diluted by purified water according to the subjects' sensitivity, to create a second confusable stimulus. The dilutions reduced the strength of the apple juice to 90–97% by weight (90%, 7 subjects; 92.5%, 6 subjects; 95%, 6 subjects; 97.5%, 1 subject). The samples were dispensed as in Experiment I at constant room temperature (20–22°C).

3.1.3. Procedure

As in Experiment I, subjects performed twelve 2-AFC and twelve 3-AFC tests in a single experimental session (session lengths ranged 14–27 min). The procedure was the same as in Experiment I except for the modifications below.

After the water rinses taken prior to each test, an apple juice sample was tasted before starting the test. This was taken as a 'primer' to prevent distortions in the flavor of the first sample of the test, if tasted after water. The primer restored the adaptation conditions of the mouth that would be encountered throughout the 2-AFC or 3-AFC. The primer was always the more dilute of the two apple juice samples.

The instructions were modified; subjects were required to pick the stronger tasting sample rather than the salt samples as in Experiment I. A warm-up was performed as before, with the difference between the two stimuli for the first warm-up being greater than for the actual test stimuli. Here, the diluted samples had 2.5–7.5% by weight lower apple juice contents.

Because the experiment was concerned only with flavor differences, it was necessary to control for the possibility of interference caused by color differences brought about by the dilution. It was noticed that color differences were perceptible when the stimuli were presented on aluminium trays, as in Experiment I.

Accordingly, the trays were painted black to obscure color differences and the experiment was performed under a masking dim red light. A control experiment in these conditions, in which the subjects went through the motions of tasting the stimuli without actually doing so, but basing their judgments on visual inspection, was performed. It indicated that subjects could not distinguish the stimuli visually (for each subject and each test, binomial $p > 0.2$).

In an initial practice session that familiarised subjects with the experiment and determined their sensitivity, six 3-AFC tests as well as six 2-AFC were performed for three sets of stimuli; For these, the dilute samples chosen were 90, 92.5 and 97% w/w apple juice.

3.1.4. Results

As in Experiment I, the proportions of correct 2-AFC and 3-AFC tests performed by each subject were used to determine appropriate d′ values (Ennis, 1993). A significant majority of subjects (15/20) obtained a higher d′ value for the 2-AFC method (binomial, $p = 0.04$). Similarly to Experiment I, one subject performed all twelve 2-AFC tests without error, and was dropped from the t-test analysis. Accordingly, mean values taken over the remaining 19 subjects were 1.2 for the 2-AFC and 0.9 for the 3-AFC. These values were significantly different (*t*-test, $p = 0.03$).

The same computation of pooled d′ values with correction for overdispersion was performed as in Experiment I. The resulting d′ values were 1.2 for the 2-AFC and 0.9 for the 3-AFC. These were found to be significantly different ($p = 0.04$) using the method of Bi et al. (1997).

The same approach to considering test power was taken as in Experiment I. For the 2-AFC, 19 tests would be required to reject H_0, while for the 3-AFC test, 30 tests would be required. Once again, the increase in d′ more than compensated for the statistical power advantage of the 3-AFC.

4. Discussion

SSA predicts that the sequences in the 2-AFC would be more favorable than in the 3-AFC procedure, yielding a higher d′ for the NaCl/water model system. There is confidence in this prediction because of the broad research on the order of discriminability of the WS, SW, WW, and SS pairs (O'Mahony & Goldstein, 1987; O'Mahony & Odbert, 1985; Tedja et al., 1994; Vié & O'Mahony, 1989). The prediction was confirmed in Experiment I. To hypothesise the same difference in d′ values for the apple juice system would be to assume the same order of discriminability of the sequences (WS, SW, WW, SS) or possibly a similar order (WS, WW, SW, SS). Both orders would predict a higher d′ for the 2-AFC. Again, the 2-AFC method elicited a higher d′ than the 3-AFC method, implying one of these orders of discriminability for the apple juice system. Either way, it can be safely stated that Rousseau and O'Mahony's (1997) findings of higher d′ values for the 2-AFC than for the 3-AFC, using yoghurt stimuli, was confirmed here for a model system and an apple juice system.

Regarding power, the increase in d′ more than compensated for the statistical power advantage of the 3-AFC, rendering the 2-AFC more powerful. This confirmed the results of Rousseau and O'Mahony, (1997).

One problem with the d′ measure, not shared by Brown's non-parametric R-index measures (Brown, 1974; O'Mahony, 1992) is the possibility of the occasional case of perfect discrimination. Here, d′ approaches infinity which would distort the *t*-test analysis used in the present study. Rousseau and O'Mahony (1997) chose an arbitrary large d′ (4.0) for such cases. Bock and Jones (1968) proposed an alternative approach. This is a correction applied to the number of tests correctly performed using the formula N(1-1/2N), where N is the number of tests performed. This approach has been called the 1/2N rule. Hautus (1995) found this rule less favorable than a log-linear rule as a result of Monte-Carlo simulations. Yet, the problem is avoided, as in the present paper, by using pooled d′ values with correction for overdispersion (Bi & Ennis, 1998).

The triadic test chosen for this study was the 3-AFC method rather than the more commonly used triangle test method. The criterion used for the 3-AFC is the same as that used for the 2-AFC method. This criterion was the degree of saltiness required for a subject to report the stimulus as being 'salt' or the intensity of apple juice flavor required to report the stimulus as being 'strong'. Such a criterion can be called a β criterion; it is the criterion used in classical Signal Detection Theory (Green & Swets, 1966). The triangle and duo-trio tests use a different type of criterion: a distance criterion. This criterion is the degree of difference required for a subject to report stimuli as being 'different'. This can be called a τ criterion. The possibility exists that in the present experiment, with the degree of stimulus learning available from repeated tasting and warm-up, that if a subject were using a triangle test, he might change from a τ and adopt a β criterion. An analysis assuming a τ criterion when a β criterion was used would overestimate d′ (Ennis, 1993). To avoid any potential problems of a change in criterion type, the 3-AFC method was chosen. It is also worth noting that the 3-AFC, being the more powerful of the two triadic methods, provides a more exacting test for the 2-AFC. Further, the version of the 3-AFC chosen used the 'strong-odd' triads (NaCl-odd, undiluted apple juice-odd) which give higher d′ values and thus greater power, again providing a more exacting test.

This paper has considered explanations based on Sequential Sensitivity analysis. However, other possible explanations should be considered. For example, it could be argued that the superior performance for the 2-AFC in Experiment I was due to greater familiarity with the test procedure, the method having been used in the practice session. This is unlikely, given that the same result was found in Experiment II, when both 2-AFC and 3-AFC methods were used in the practice session. Also, the effect of practice is most likely to be stimulus learning rather than favouring one of two relatively simple test protocols (Tedja et al., 1994).

Another way of expressing the results would be to say that the variance of the stimuli in the 3-AFC condition is greater than in the 2-AFC condition. Certainly, the sequencing of the stimuli would cause the sensations elicited by NaCl or distilled water to be changed more because of interactions between the stimulus and the conditions in the mouth. This would increase the variance.

With stimuli like food or tastants, which interact such that stimulus perception can be affected by preceding stimuli, Thurstonian considerations must be modified. Ennis and O'Mahony (1995) approached this problem by regarding a given stimulus, S, not as a single stimulus, but considering S tasted after W (W–S) and S tasted after S (S–S) as separate stimuli. Thus, the sequence of tasting in which S occured would have an effect on the sensation elicited by S. In different sequences (W–S vs S–S) the sensations elicited by this stimulus would be sufficiently different that the brain could regard them as having been elicited by separate stimuli. One stimulus would be S conditional on a priorly tasted W stimulus, and another, eliciting a perceptually separate sensation, would be S conditional on a priorly tasted S stimulus. In the same way, W could also be considered as two stimuli (W–W) and (S–W). Ennis and O'Mahony (1995) introduced the concept of a conditional stimulus to describe such stimuli that are created partially by the effects of the oral environment. Taking this approach for Tedja et al.'s (1994) data with the model NaCl/water system, a four distribution Thurstonian model was developed as an alternative to the traditional two distribution Thurstone-Ura approach. It successfully modelled Tedja et al.'s results concerning differences in sequencing for triadic tests. The same conditional stimulus model, combining the elements of the Thurston-Ura and SSA approaches would also predict a higher d' for the 2-AFC than for the 3-AFC method.

There is one further consideration regarding the Thurstonian modelling approach applied here. Simple Thurstonian models consider distributions along a univariate axis, usually of intensity. So traditional instructions for the 2-AFC test would be to pick the more intense or stronger of 2 stimuli. For the 3-AFC, they would be to tell the subjects that of three stimuli one was stronger and two were weaker and to pick the stronger stimulus (or vice versa). Such instructions are adequate when stimulus intensity is the only cue that varies with increasing physical strength. However, this is not always the case in the chemical senses. With increasing concentration, taste stimuli may not only increase in perceived intensity but may also change in taste quality. Subjects report a variety of different tastes as the concentration of NaCl increases (O'Mahony, 1973; O'Mahony et al., 1976). This phenomenon is a gustatory analogue of the Bezold–Brücke phenomenon. To take a visual example, a red light as it becomes more intense, also changes hue. This is called the Bezold–Brücke phenomenon. However, judgements of intensity of the red light are entirely possible in spite of this hue change. The subject merely allows for the fact that the visual intensity axis includes a certain degree of hue change along with intensity change. The same is true for increasing concentrations of NaCl in distilled water. Distilled water itself produces a sensation on the tongue which slowly changes as NaCl is added. This changed sensation gradually becomes stronger until it has a quality that most people would call 'salty'. However, at lower concentrations the various tastes elicited by NaCl solutions would not be generally described as 'salty'. Unless subjects were familiar with these sensations, they might confuse the taste of distilled water with that of low concentration NaCl. To use the traditional 2-AFC or 3-AFC instructions by merely asking subjects to identify the stronger or more intense stimulus, would invite such errors. Hence the instructions were given in the present study did not require subjects merely to consider intensity. After using warm-up to identify the sensations elicited by distilled water and NaCl, the instructions were modified for the 2-AFC and 3-AFC tests to identify the salt stimulus. In doing this however, subjective reports indicated that subjects approached the task by deciding which stimuli were salt and which were water. This could be visualised in Thurstonian terms by considering distributions along a univariate intensity axis, which also included taste quality change, as the intensity increased. At one end of the axis would be the distilled water taste and at the other end of the axis the taste quality associated with the strongest NaCl tastes encountered during the experiment. Subjects found the task simple and could assess changes in NaCl stimulus strength taking into account both intensity and quality change.

The argument above visualises distributions along a single axis which represents the intensity change incorporating taste quality change. Yet, this is not the only way the data could be modelled. The stimulus sensation for water and NaCl could be assessed in terms of their proximity to stereotypes for water and NaCl. Stimuli closer to the NaCl stereotype would be categorised as NaCl while stimuli closer to the water stereotype would be categorised as water. The stereotypes themselves being memory engrams would also have a variance which might well be larger than the stimulus variances. Taking this approach and also considering inequality of these variances, it is possible to account for the present results. Although subjective reports would indicate the potency of sequence effects considered by the SSA model, an alternative Thurstonian approach may well be found to account at least in part for the differences in d′ between the 2-AFC and the 3-AFC methods. Certainly more research is required to distinguish between these various Thurstonian approaches.

References

Bi, J., & Ennis, D. M. (1998). A Thurstonian variant of the beta-binomial model for replicated difference tests. Submitted.

Bi, J., Ennis, D. M., & O'Mahony, M. (1997). How to estimate the variance of d′ from difference tests. *Journal of Sensory Studies, 12*, 87–104.

Bock, R. D., & Jones, L. V. (1968). *The measurement and prediction of judgement and choice*. San Francisco: Holen-Day.

Brown, J. (1974). Recognition assesed by rating and ranking. *British Journal of Psychology, 65*, 13–22.

Buchanan, B., Givon, M., & Goldman, A. (1987). Mesurement of discrimination ability in taste tests: an empirical investigation. *Journal of Marketing Research, 24*, 154–163.

Byer, A. J., & Abrams, D. (1953). A comparison of the triangular and two-samples taste-test methods. *Food Technology, 7*, 185–187.

Dawson, E. H., & Dochterman, E. (1951). A comparison of sensory methods of measuring differences in food qualities. *Food Technology, 5*, 79–81.

Ennis, D. M. (1988a). Multivariate sensory analysis. *Food Technology, 42*, 118–122.

Ennis, D. M. (1988b). Confusable and discriminable stimuli: Comment on Nosofsky (1986). & Shepard (1986). *Journal of Experimental Psychology: General, 117*, 408–411.

Ennis, D. M. (1990). Relative power of difference testing methods in sensory evaluation. *Food Technology, 44*, 114–118.

Ennis, D. M. (1992). Modeling similarity and identification when there are momentary fluctuations in psychological magnitude. In: F. G. Ashby (Ed.), *Multidimensional models of perception & cognition*, (pp. 279–298). Hillsdale, New Jersey: Lawrence Erlbaum Assoc.

Ennis, D. M. (1993). The power of sensory discrimination methods. *Journal of Sensory Studies, 8*, 353–370.

Ennis, D. M., & Bi, J. (1998). The beta-binomial model: accounting for inter-trial variation in replicated difference and preference tests. Submitted.

Ennis, D. M., & Mullen, K. (1985). The effect of dimensionality on results from the triangular method. *Chemical Senses, 10*, 605–608.

Ennis, D. M., & Mullen, K. (1986a). Theoretical aspects of sensory discrimination. *Chemical Senses, 11*, 513–522.

Ennis, D. M., & Mullen, K. (1986b). A multivariate model for discrimination methods. *Journal of Mathematical Psychology, 30*, 206–219.

Ennis, D. M., & Mullen, K. (1992a). Probabilistic psychophysics with noisy stimuli. *Mathematical Social Sciences, 23*, 221–234.

Ennis, D. M., & Mullen, K. (1992b). A general probabilistic model for triad discrimination, preferential choice, and two-alternative identification. In: F. G. Ashby (Ed.), *Multidimensional models of perception & cognition*, (pp. 115–122). Hillsdale, New Jersey: Lawrence Erlbaum Assoc.

Ennis, D. M., & O'Mahony, M. (1995). Probabilistic model for sequential taste effects in triadic choice. *Journal of Experimental Psychology: Human Perception & Performance, 21*, 1088–1097.

Ennis, D. M., Mullen, K., & Frijters, J. E. R. (1988). Variants of the method of triads: Unidimensional Thurstonian models. *British Journal of Mathematical & Statistical Psychology, 41*, 25–36.

Ennis, D. M., Palen, J. J., & Mullen, K. (1988). A multidimensional stochastic theory of similarity. *Journal of Mathematical Psychology, 32*, 449–465.

Filipello, F. (1956). A critical comparison of the two-sample and triangular binomial designs. *Food Research, 21*, 235–241.

Francois, P., & Sauvageot, F. (1988). Comparison of the efficiency of pair, duo-trio and triangle tests. *Journal of Sensory Studies, 3*, 81–94.

Frijters, J. E. R. (1979a). The paradox of discriminatory nondiscriminators resolved. *Chemical Senses & Flavor, 4*, 355–358.

Frijters, J. E. R. (1979b). Variations of the triangular method and the relationship of its unidimensional probabilistic models to three-alternative forced choice signal detection theory models. *British Journal of Mathematical & Statistical Psychology, 32*, 229–241.

Frijters, J. E. R. (1980). Three-stimulus procedures in olfactory psychophysics: an experimental comparison of Thurstone-Ura and three-alternative forced choice models of signal detection theory. *Perception & Psychophysics, 28*, 390–397.

Frijters, J. E. R. (1981a). An olfactory investigation of the compatibility of oddity instructions with the design of a 3-AFC signal detection task. *Acta Psychologica, 49*, 1–16.

Frijters, J. E. R. (1981b). The excitatory state in the triangular constant method. *Psychometrika, 46*, 219–222.

Frijters, J. E. R. (1982). Expanded tables for conversion of a proportion correct responses (P_c) to the measure of sensory difference (d′) for the triangular method and the 3-alternative forced choice procedure. *Journal of Food Science, 47*, 139–142.

Frijters, J. E. R., Kooistra, A., & Vereijken, P. F. G. (1980). Tables of d′ for the triangular method and the 3-AFC signal detection procedure. *Perception & Psychophysics, 27*, 176–178.

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: John Wiley & Sons.

Gridgeman, N. T. (1955). Taste comparisons: two samples or three?. *Food Technology, 9*, 148–150.

Gridgeman, N. T. (1956). Group size in taste sorting trials. *Food Research, 21*, 534–539.

Gridgeman, N. T. (1970). A re-examination of the two-stage triangle test for the perception of sensory differences. *Journal of Food Science, 35*, 87–91.

Grim, A. C., & Goldblith, S. (1965). Some observed discrepancies in application of the triangle test to evaluation of irradiated whole-egg magma. *Food Technology, 19*, 146.

Hacker, M. J., & Ratcliffe, R. (1979). A revised table of d′ for M-alternative forced choice. *Perception & Psychophysics, 26*, 168–170.

Hautus, M. J. (1995). Correction for extreme proportions and their biasing effects on estimated values of d′. *Behavioral Research Methods, Instruments & Computers, 27*, 46–51.

Helm, E., & Trolle, B. (1946). Selection of a taste panel. *Wallerstein Lab. Communications, 9*, 181–194.

Hopkins, J. W. (1954). Some observations on sensitivity and repeatability of triad taste difference tests. *Biometrics, 10*, 521–530.

Hopkins, J. W., & Gridgeman, N. T. (1955). Comparative sensitivity of pair and triad flavor intensity difference tests. *Biometrics, 11*, 63–68.

Kapenga, J. A., de Doncker, E., Mullen, K., & Ennis, D. M. (1987). The integration of the multivariate normal density function for the triangular method. In: P. Keast, G. Fairweather (Eds.), *Numerical Integration*, (pp. 321–328.) Dordrecht: D. Reidel Publ. Co.

MacMillan, N. A., & Creelman, C. D. (1991). *Detection Theory: A User's Guide*. Cambridge: Cambridge University Press.

MacRae, A. W., & Geelhoed, E. N. (1992). Preference can be more powerful than detection of oddity as a test of discriminability. *Perception & Psychophysics, 51*, 179–181.

Masuoka, S., Hatjopoulos, D., & O'Mahony, M. (1995). Beer bitterness detection: testing Thurstonian and sequential sensitivity analysis models for triad and tetrad methods. *Journal of Sensory Studies, 10*, 295–306.

Mosteller, F. (1951a). Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika, 16*, 3–9.

Mosteller, F. (1951b). 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.

Mosteller, F. (1951c). 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.

Mullen, K., & Ennis, D. M. (1987). Mathematical formulation of multivariate Euclidiean models for discrimination methods. *Psychometrika, 52*, 235–249.

Mullen, K., & Ennis, D. M. (1991). A simple multivariate probabilistic model for preferential and triadic choices. *Psychometrika, 56*, 69–75.

Mullen, K., Ennis, D. M., de Doncker, E., & Kapenga, J. A. (1988). Models for the duo-trio and the triangular methods. *Biometrics, 44*, 1169–1175.

O'Mahony, M. (1973). Qualitative description of low concentration sodium chloride solutions. *British Journal of Psychology, 64*, 601–606.

O'Mahony, M. (1979). 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.

O'Mahony, M. (1986). Sensory adaptation. *Journal of Sensory Studies, 1*, 237–258.

O'Mahony, M. (1992). Understanding discrimination tests: A user-friendly treatment of response bias, rating and ranking R-index tests and their relationship to signal detection. *Journal of Sensory Studies, 7*, 1–47.

O'Mahony, M. (1995). Who told you the triangle test was simple?. *Food Quality & Preference, 6*, 227–238.

O'Mahony, M., & Goldstein, L. (1986). Effectiveness of sensory difference tests: Sequential Sensitivity Analysis for liquid food stimuli. *Journal of Food Science, 51*, 1550–1553.

O'Mahony, M., & Goldstein, L. (1987). Tasting successive salt and water stimuli: The roles of adaptation, variability in physical signal strength, learning, supra- and subadapting signal detectability. *Chemical Senses, 12*, 425–436.

O'Mahony, M., Kingsley, L., Harji, M., & Davies, M. (1976). What sensation signals the salt taste threshold? *Chemical Senses and Flavor, 2*, 177–188.

O'Mahony, M., Masuoka, S., & Ishii, R. (1994). A theoretical note on difference tests: Models, paradoxes and cognitive strategies. *Journal of Sensory Studies, 9*, 247–272.

O'Mahony, M., & Odbert, N. (1985). A comparison of sensory difference procedures: Sequential Sensitivity Analysis and aspects of taste adaptation. *Journal of Food Science, 50*, 1055–1058.

O'Mahony, M., Thieme, U., & Goldstein, L. (1988). The warm-up effect as a means of increasing the discriminability of sensory difference tests. *Journal of Food Science, 53*, 1848–1850.

O'Mahony, M., Wong, S. Y., & Odbert, N. (1986). Sensory difference tests: Some rethinking concerning the general rule that more sensitive tests use fewer stimuli. *Lebensmittel -Wissenschaft und-Technologie, 19*, 93–94.

Peryam, D. R. (1958). Sensory difference tests. *Food Technology, 12*, 231–236.

Peyram, D. R., & Swartz, V. W. (1950). Measurement of sensory differences. *Food Technology, 4*, 390–395.

Pokorny, J., Marcin, A., & Davidek, J. (1981). Comparison of the efficiency of triangle and tetrade tests for discriminatory sensory analysis of food. *Die Nahrung, 25*, 561–564.

Pfaffmann, C. (1954). Variables affecting difference tests. In: D. R. Peryam, F. S. Pilgrim, & M. S. Peterson (Eds.), *Food acceptance testing methodology: A symposium*, Washington, DC.: Natl. Acad. Sci. - Nat. Res. Council.

Rousseau, B., & O'Mahony, M. (1997). Sensory difference tests: Thurstonian and SSA predictions for flavored yogurts. *Journal of Sensory Studies, 12*, 127–146.

Stillman, J. A. (1993). Response selection, sensitivity, and taste-test performance. *Perception & Psychophysics, 54*, 190–194.

Tedja, S., Nonaka, R., Ennis, D. M., & O'Mahony, M. (1994). Triadic discrimination testing: Refinement of Thurstonian and Sequential Sensitivity Analysis approaches. *Chemical Senses, 19*, 279–301.

Thieme, U., & O'Mahony, M. (1990). Modifications to sensory test protocols: the warmed up paired comparison, the single standard duo-trio and the A-not A test modified for response bias. *Journal of Sensory Studies, 5*, 159–176.

Thurstone, L. L. (1927a). Psychophysical analysis. *American Journal of Psychology, 38*, 368–389.

Thurstone, L. L. (1927b). A law of comparative judgement. *Psychological Review, 34*, 273–286.

Ura, S. (1960). Pair, triangle and duo-trio test. *Reports of Statistical Application Research, Japanese Union of Scientists & Engineers, 7*, 107–176.

Vessereau, A. (1965). Les méthodes statistiques appliquées au test des caractères organoleptiques. *Annales de Nutrition et d'Alimentation, 19*, 103–140.

Vié, A., & O'Mahony, M. (1989). Triangular difference testing: Refinements to Sequential Sensitivity Analysis for predictions for individual triads. *Journal of Sensory Studies, 4*, 87–103.

Wasserman, A. E., & Talley, F. (1969). A sample bias in the evaluation of smoked frankfurters by the triangle test. *Journal of Food Science, 34*, 99–100.