Response selection, sensitivity, and taste-test performance

J. A. STILLMAN

University of Auckland, Auckland, New Zealand

Tasters selected the odd stimulus from among sets of three samples of party dip. Two samples came from one batch, and one sample came from another batch. The physicochemical difference between the batches consisted of the presence or absence of added salt. Two different tests of discriminability were undertaken by the same subjects with the same stimuli: the triangle test and the three-alternative forced-choice (3-AFC) method. Although different numbers of correct selections were obtained in the two tasks, an index of discriminability, d' , had the same value when the data were analyzed in accordance with the Thurstone-Ura and signal-detection models, respectively. The average data support Frijters's (1979b) contention that different models of the discrimination process are appropriate to the results of the triangular and the 3-AFC procedures. Further analysis of the data revealed that discrimination was poorer for trios containing one physicochemically weak stimulus and two stronger stimuli than it was for trios containing one stronger stimulus and two weak stimuli. A two-signal 3-AFC task was undertaken by some subjects, and d' estimates from this task were lower than expected on the basis of performance in the other tasks.

In food science, seemingly paradoxical results are obtained when two superficially similar tests of discriminability, each employing tristimulus sets containing one "odd" sample, are administered to a group of tasters. In a three-alternative forced-choice (3-AFC) signal-detection task, the sensory attribute correlated with the physicochemical difference between the samples is specified, and the percentage of correct choices of the odd stimulus is greater than it is with the use of the triangular method (Amerine, Pangborn, & Roessler, 1965), in which the difference is not specified and tasters are required merely to identify the odd stimulus among three. This result appears surprising, since, to quote Gridgeman (1970), "It is natural to assume that a person who can't identify the oddity ipso facto can't discriminate, much less characterize, the difference between the two products."

Frijters (1979b) argued that different models of the discrimination process are appropriate to each of the tasks. Both models assume that samples from the two sources, say A and B, each give rise to internal sensory responses that, in the simplest case, are normally distributed with equal variances. In either model, therefore, the distance between the means of the two distributions, in units of

their common standard deviation, can serve as an index of discriminability. This index is the d' of detection theory (Green & Swets, 1966). Thus, if each model is valid, estimates of d' extracted from the results of each task should not differ significantly, in spite of the expectation of disparate numbers of percent correct choices in the two tasks (Frijters, Kooistra, & Vereijken, 1980).

The Thurstone-Ura model (Frijters, 1979a) is applicable to the triangle test. In this test, the subject is required to select the odd stimulus from each tristimulus set, and it is assumed that a correct response will be made if the difference between the momentary sensory values corresponding to the two samples from the same source is less than the difference between each of these and the momentary sensory value corresponding to the remaining sample. Given this decision rule, the probability of a correct response in the triangular method is

$$P_c = 2 \int_0^\infty \left\{ \Phi[-u\sqrt{3} + d'\sqrt{(2/3)}] + \Phi[-u\sqrt{3} - d'\sqrt{(2/3)}] \right\} \frac{e^{-1/2u^2}}{\sqrt{2\pi}} du, \quad (1)$$

where $\Phi(u)$ is the cumulative standard normal distribution, $d' = (\mu_B - \mu_A)/\sigma$, and $\mu_A < \mu_B$.

In the 3-AFC procedure, the subject must select the strongest stimulus with respect to some attribute from a tristimulus set containing two samples from the physicochemically weaker source. Under these instructions, a correct response is assumed to be made if the momentary sensory value associated with the odd sample is greater than that associated with the two samples from the other source. The probability of being correct in the 3-AFC task is the

The author gratefully acknowledges the assistance of Sze Tan, of the Department of Physics, University of Auckland, who was responsible for evaluating the integral in Equation 2 and for creating a computer program to output either d' or percent correct values for the one- and two-signal three-alternative forced-choice tasks. Thanks are also due R. J. Irwin, who gave useful comments on a draft of this article, and Hansells Ltd (N.Z.), who donated the dip mix. Correspondence should be addressed to the author at the Department of Psychology, Massey University Albany, Private Bag 102-904, North Shore MSC, Auckland, New Zealand.

probability that the sensory value associated with the physicochemically different sample is greater than the maximum value of the other two samples. This probability is

$$P_c = \int_0^{\infty} [\Phi^2(u+d') + \Phi^2(-u+d')] \frac{e^{-\frac{1}{2}u^2}}{\sqrt{2\pi}} du. \quad (2)$$

An analogous result is obtained in the case in which the subject must select the weakest stimulus from a tristimulus set containing two samples from the physicochemically stronger source.

In summary, under the instruction to select the odd man out, distances between momentary sensory values are compared, whereas under the instruction to select the strongest (or weakest) stimulus, the subject selects the stimulus correlated with the greatest (or least) momentary sensory value, irrespective of the distance between this sensation and that associated with the other two samples.

Recently two papers have applied Frijters's (1979a, 1979b) analysis to data collected for other purposes. In each case, the data to which the models have been applied have been obtained in experiments with features that make them less than ideal tests of the models. The present paper reports an intentional application of the analysis proposed by Frijters. First, however, I discuss the analyses already referred to. Frijters recognized that the so-called two-stage triangle test (Gridgeman, 1970) amounts to the application of a triangle test followed by a 3-AFC procedure. He therefore applied the models to the data from Experiment 2 of Byer and Abrams (1953), who used this procedure to study the discriminability of weak compounds. Byer and Abrams required tasters to select the odd sample from three glasses, two containing the same quinine solution and one containing a different quinine solution. The tasters were then required to state whether they considered the sample selected as odd to be more or less bitter than the two samples judged as being alike. Although more than half the panelists (53%) misidentified the odd sample of quinine sulfate solutions, most of them (70.8%) correctly identified the weaker or stronger solution when they were asked to select the most (or least) bitter sample of the three. Frijters (1979a) pointed out that in Byer and Abram's study percent correct in the initial triangle task (46.67) was equal to a d' of 1.29, and percent correct in the 3-AFC stage was equal to a d' of 1.28. The concordance of these d' estimates is impressive, although the 3-AFC procedure was not used explicitly by Byer and Abram, and when results from the second stage of the experiment were scored, responses of tasters who had failed to identify the odd sample were taken into account under the assumption that the two samples not selected gave rise to indistinguishable sensory responses corresponding to the expected sensory value of the actual odd sample.

In a study of preference, MacRae and Geelhoed (1990, 1992) presented tasters with two samples of distilled water and one sample of tap water. They found that tasters were significantly more consistent in choosing the tap water as most preferred than they were in identifying it as the odd sample. These authors then applied Frijters's analysis to

their data and obtained similar estimates of d' from the two tasks (d' triangle = 1.62, d' 3-AFC = 1.31). However, preference is not an ideal test of the models, because preferences cannot be either correct or incorrect, and in order to perform the analysis, it was necessary to assume that the preference was strongly in favor of tap water. The study did not consider tristimulus sets containing two samples of tap water and one sample of distilled water. However, in a similar study, Geelhoed and MacRae (1991) included sets containing two samples of tap water.

The present experiment provides a specific test of Frijters's (1979b) prediction that percent correct choices for the same sets of stimuli obtained by the triangular method and by the 3-AFC method will yield a comparable estimate of d' if the results are analyzed in accordance with the Thurstone-Ura and signal-detection models, respectively. The effect of different stimulus arrangements is also investigated.

METHOD

Subjects

The subjects were 144 unpaid volunteers who were undergraduates at the University of Auckland.

Stimuli

The stimuli were samples of onion party dip (provided by Hannels Ltd of New Zealand) made up in accordance with the manufacturer's instructions, except as described below. This product was chosen because it allows good experimental control. It consists of a fine powder that is mixed with milk. Half of the samples were made up from powder to which salt was added at the rate of 2 g per 215 g of powder. Seven hundred milliliters of pasteurized homogenized milk was mixed with each 215 g of powder. The powder had been sifted to remove the onion flakes usually incorporated in the product. The mixes were prepared in the evening and refrigerated for use the following day. Samples were presented to the tasters at room temperature, approximately 21°C.

Procedure

For each of the two tasks—the triangle task and the 3-AFC task—all 6 possible temporal (and spatial) orders of the samples from the two sources were used. If N designates the batch with no added salt, and S the batch with added salt, the possible arrangements are SSN, SNS, NSS, NNS, NSN, and SNN. Each taster completed the triangle test first, remaining unaware of the physicochemical difference between batches. Each of the 6 orders in the triangle task was paired with all 6 orders in the 3-AFC task; therefore, there were 36 possible arrangements of the 6 samples judged by each taster in the course of the experiment. In the first part of the experiment, these 36 orders were used three times each with a group of 108 tasters. In the second part of the experiment, the 36 orders were each used once more by a separate group of 36 individuals. In the 3-AFC task, the latter group made a choice of either least salty or most salty from a set containing two samples from the appropriate batch. In signal-detection terms, a 3-AFC trial for members of this group comprised two signal intervals and one noise interval. Samples were arranged on two trays, one for each task, in rows of three identical plastic pots. Both general verbal instructions and specific written instructions were given. The written instructions were on a folded sheet that was turned over when the first task was completed. Instructions referred the subject to a designated row for each task, and each sample was obtained with a different plastic spoon. After each taste, the spoons were placed in order on a paper napkin, with a residue remaining to allow a sec-

ond taste if required. Tasting and retasting were always in a row from left to right. The tasters registered their choices by marking one of three circles drawn on the instruction sheet to represent the positions of the samples in a row. Included among the instructions for the triangle task was the following: "Taste the three samples from left to right. When you have tasted them all, please pick out one sample that seems most different from the other two." Two types of instruction sheet were printed so that the 3-AFC task tasters who were presented with two N-samples and one S-sample received an instruction to select the one sample that seemed "most salty," whereas tasters who were presented with two S-samples and one N-sample received an instruction to select the sample that seemed "least salty."

RESULTS

The average outcome from the first group of 108 tasters is in line with predictions for the two models of taste discrimination discussed by Frijters (1979a). In the triangle task, 42/108 (38.89%) correct selections of the odd stimulus were made. This corresponds to a d' of 0.80. In the 3-AFC task, 62/108 (57.41%) correct selections of the odd stimulus were made. This also corresponds to a d' of 0.80 (d' estimates to two decimal points). It is evident that a slightly stronger concentration of salt would have been advantageous. In pilot work involving 43 tasters, sets comprising one extra-salt and two no-added-salt trios were used. For these trials, the extra-salt batch was made up with 1.5 g of salt added to each 215 g of dip powder, and the percent correct was 37% ($d' = 0.64$). Consequently, the amount of salt to be used in the experiment was increased by 33% to 2 g, and the percent correct for SSN trios then increased to 46.30%. After data were collected, however, it became apparent that discrimination might not be constant across trio types, and that performance with SSN trios might be better than performance with NNS trios. This is discussed below.

The probability of a correct response in the two-signal 3-AFC task undertaken by the separate group of 36 tasters is the probability that either of the signals is identified correctly. When the two signal samples have the stronger concentration of salt, the probability is one minus the probability that the sensory value associated with the odd sample is greater than the maximum value of the two other samples, or in other words, one minus the integral in Equation 1, except that $-d'$ replaces d' . A complementary situation exists when the two signal samples have the weaker concentration of salt. Results for the tasters who completed the two-signal 3-AFC task were as follows: Percent correct in the triangle task = 38.89 (14/36), giving a d' of 0.8, which was identical to the value obtained by Group 1. Percent correct in the two-signal 3-AFC task = 77.80 (28/36), which was equivalent to a d' of 0.43. The chance expectation in the two-signal case was 66.67%, and 85.50% correct selections would have been required to equal the d' of 0.8 attained in the triangle task.

A finer grained examination of the results is warranted, because although overall the results were precisely in line with expectations, inspection of the data suggests that performance decreased for trios containing two samples from the batch without the added salt. Table 1 gives the num-

Table 1
Analysis of Frequencies of Choice of Samples With and Without Added Salt Under Different Stimulus Arrangements in Three Tasks

Task	Choice	Trio Type	Number Correct	Percent Correct	d'
Group 1					
Triangle	Odd man out	NNS	25/54	46.30	1.25
	Odd man out	SSN	17/54	31.48	0
One-signal 3-AFC	Most salty	NNS	34/54	62.96	0.99
	Least salty	SSN	28/54	51.85	0.62
Group 2					
Triangle	Odd man out	NNS	7/18	38.90	0.80
	Odd man out	SSN	7/18	38.90	0.80
Two-signal 3-AFC	Most salty	NNS	14/18	77.80	0.43
	Least salty	SSN	14/18	77.80	0.43

Note—Estimates of d' are also given. Trios contained two samples with added salt (S) or two samples with no added salt (N).

ber correct, percent correct, and d' estimates for the two groups in each type of task and stimulus grouping.

For Group 1, the proportion of correct choices in both the triangle and the 3-AFC tasks was smaller with SSN trios than with NNS trios, and this reduction is reflected in the index of sensitivity, d' . No effect of trio type was evident for the subjects in Group 2, although the number of subjects tasting SSN and NNS trios was too small to be sensitive to a difference of the magnitude demonstrated by Group 1 (15%).

DISCUSSION

For the tasters in Group 1, performance in both the triangle and the 3-AFC tasks was poorer when the odd sample was from the batch that did not contain added salt than when it was from the batch that did contain added salt. It may be that trios containing one weaker stimulus on the relevant sensory dimension provide a more difficult discrimination than do trios containing one stronger stimulus. However, the performance decrement was especially marked in the triangle task, in which, with this combination, performance was just below chance expectation. No decrement with this combination was evident for the second, admittedly smaller, group of tasters. On inspection, the extent of the decrement for trios containing two added-salt samples is exaggerated by below-chance judgments on the part of one group of 18 tasters in the triangle task—namely, those for whom the temporal ordering was SSN. In both tasks, all other permutations led to performance at greater than chance levels. In the light of the possibility that trios containing one weaker stimulus on the relevant sensory dimension provide a more difficult discrimination, the data from tasks A, C, and D in Table 1 of Geelhoed and MacRae (1991) were examined, and d' estimates were calculated from them. The results of the present experiment were also compared with theirs with respect to the decrement evident for two-signal 3-AFC trios.

In Table 2, the data from Table 1 of Geelhoed and MacRae (1991) have been rearranged to make explicit the

two groups of subjects created by the authors' counterbalancing procedure. In the analysis given in Table 2, T represents a sample of tap water, and D represents a sample of distilled water. Geelhoed and MacRae's contention that tap water was preferred by a large majority of subjects is accepted, and it is assumed that tap water had more of the sensory attribute underlying this preference than did distilled water. There is no evidence of a decrement in discrimination for TTD trios in Geelhoed and MacRae's triangle data; however, such a decrement is suggested in their one-signal 3-AFC data. It may be, therefore, that for moderately confusable stimuli, a single strong stimulus "stands out" perceptually in the context of two weaker samples more than does a single weak stimulus in the context of two stronger samples. It is interesting to note that analogous phenomena have been observed in another sensory system, audition, in which intensity increments in a sound are detected better than intensity decrements (Macmillan, 1971), and brief tone bursts are more readily detectable than brief gaps of equal duration (Irwin & Kemp, 1976).

In both the present experiment and the one conducted by Geelhoed and MacRae (1991), performance in the two-signal 3-AFC task does not exceed chance to the extent expected on the basis of performance in the one-signal 3-AFC task. The reason for this outcome is not obvious and requires further investigation. In this experiment, the discrimination was not easy, and the subjects in the two-signal 3-AFC task were not told that two samples out of the three were from the source correlated with the dimension that they were to judge. The tasters were merely asked to pick the one sample that *seemed* most (or least) salty. If the presumption was that just one of the three samples was likely to produce the requisite experience, then on occasions when the physicochemically odd stimulus produced the most distinct momentary sensory experience, it may have been selected and consequently mislabeled as, for example, "most salty" rather than "least salty."

In Figure 1, the average outcome of the two tasks in the first part of this experiment is shown alongside the results of three other experiments—namely, those of MacRae and

Figure 1. Percent correct selections of the odd stimulus from a set of three, as a function of d' . Data are from this study (open squares = saltiness) and from studies by MacRae and Geelhoed (1992; closed circles = preference), Geelhoed and MacRae (1991; crossed = preference), and Byer and Abrams (1953; open circles = bitterness). The solid lines are the theoretical functions relating the probability of a correct response and values of d' for the triangular and three-alternative forced-choice models.

Geelhoed (1992) and Geelhoed and MacRae (1991), who studied preference, and Byer and Abrams (1953), who studied the perception of bitterness. The solid lines are the theoretical functions relating the probability of a correct response to values of d' for the triangular and 3-AFC models. An exact agreement with the models requires that pairs of points from each of the studies share a common abscissa. In previous analyses undertaken in support of Frieters's resolution of apparently paradoxical results in taste research, no account had been taken of particular stimulus arrangements. However, because all possible arrangements are normally used within a study (MacRae & Geelhoed, 1992, is an exception), the analysis of average data provides an adequate test of his proposal that different models of the discrimination process are appropriate to the triangular and the 3-AFC procedures. Clearly, the data in Figure 1 are in good accord with expectations based on that proposal.

REFERENCES

- AMERINE, M. A., PANGBORN, R. M., & ROESSLER, E. B. (1965). *Principles of sensory evaluation of food*. New York: Academic Press.
- BYER, A. J., & ABRAMS, D. (1953). A comparison of the triangular and two-sample taste-test methods. *Food Technology*, 7, 185-187.
- FRIJTERS, J. E. R. (1979a). The paradox of discriminatory nondiscriminators resolved. *Chemical Senses & Flavour*, 4, 355-359.
- 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., KOOISTRA, A., & VEREIJEN, P. F. G. (1980). Tables of d' for the triangular method and the 3-AFC signal detection procedure. *Perception & Psychophysics*, 27, 176-178.
- GEEHOED, E. N., & MACRAE, A. W. (1991, October). An eternal triangle: Discrimination, preference, choice. In G. R. Lockhead (Ed.), *Fechner Day 91: Proceedings of the 7th Annual Meeting of the International Society for Psychophysics* (pp. 61-65). Duke University.

Table 2
Analysis of Frequencies of Choice of Tap and Distilled Water in Different Tasks

Task	Choice	Trio Type	Number Correct	Percent Correct	d'
Group 1					
Triangle	Odd man out	DDT	20/48	41.67	1.01
One-signal 3-AFC	Most preferred	DDT	32/48	66.67	1.13
	Least preferred	TTD	28/48	58.33	0.82
Group 2					
Triangle	Odd man out	TTD	20/48	41.67	1.01
Two-signal 3-AFC	Most preferred	TTD	35/48	72.92	0.23
	Least preferred	DDT	35/48	72.92	0.23

Note—Data are reordered from Table 1 of an experiment by Geelhoed and MacRae (1991). Estimates of d' are also given. Trios contained two samples of tap water (T) or two samples of distilled water (D).

- GREEN, D. M., & SWETS, J. A. (Eds.) (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- GRIDGEMAN, N. T. (1970). A reexamination of the two-stage triangle test for the perception of sensory differences. *Journal of Food Science*, **35**, 89-91.
- IRWIN, R. J., & KEMP, S. (1976). Temporal summation and decay in hearing. *Journal of the Acoustical Society America*, **59**, 920-925.
- MACMILLAN, N. A. (1971). Detection and recognition of increments and decrements in auditory intensity. *Perception & Psychophysics*, **10**, 233-238.
- MACRAE, A. W., & GEELHOED, E. N. (1990). Sensory preference beats sensory discrimination. In F. Müller (Ed.), *Fechner Day 90: Proceedings of the 6th Annual Meeting of the International Society for Psychophysics* (pp. 246-250), Würzburg, Germany.
- 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.

(Manuscript received May 19, 1992;
revision accepted for publication January 20, 1993.)