

Taste-quality recognition and forced-choice response

JAMES M. WEIFFENBACH

National Institute of Dental Research, National Institutes of Health, Bethesda, Maryland 20205

The Collings (1974) procedure for measuring taste-quality recognition thresholds controls adequately for bias caused by variation in the willingness of subjects to report perceiving a taste quality. However, it fails to control for bias due to the preference of subjects for one response over another. In the present study, preferential selection among the response alternatives was reflected in the disproportionate occurrence of the four taste-quality names among incorrect responses. Response preferences were idiosyncratic and grew stronger with repeated testing. They were not eliminated by informing subjects of the correctness of their responses. Use of this procedure for measuring recognition thresholds is discouraged.

The taste-quality recognition threshold is the least concentration of a substance that elicits the subjective experience of its characteristic taste quality. Threshold measurement procedures usually require subjects to report whether or not individual samples of fluid elicit a specified taste quality. Under these conditions, thresholds reflect not only the subjects' taste sensitivity but also their criterion for deciding that they perceived a taste quality. Because subjects are given the option of reporting that no quality was perceived, their willingness to report experiencing a taste quality affects the threshold. The tendency to report a subjective experience when the adequate stimulus for it is absent may be assessed and perhaps be manipulated by occasionally presenting blank stimuli. However, there appears to be no way to control for the opposite tendency, the tendency not to report a taste quality when one is actually experienced. As an alternative to accepting reports of subjective experience, an investigator can require subjects to demonstrate objectively that they have perceived some aspect of the stimulus.

Forced-choice procedures (Jones, 1956; Blackwell, Note 1) provide a general paradigm for requiring subjects to demonstrate their sensitivity. One way to demonstrate that subjects have extracted information from taste stimuli is to require them to separate some (usually four) samples of a taste solution from an equal number of samples of water (Harris & Kalmus, 1949). Thresholds obtained by this procedure are detection thresholds. Thresholds have also

been obtained under modified instructions requiring the subjects to separate the samples on the basis of a specified taste quality (Wotman, Mandel, Thompson, & Lavagh, 1967; Yasaki, Miyashita, Ahiko, Hirano, Kamata, & Iizuka, 1976; Zengo & Mandel, 1972). The interpretation of such thresholds is ambiguous because samples that elicit no recognizable taste quality may nonetheless be discriminated from tasteless ones. However, it is clear that thresholds obtained with the modified procedure are not necessarily recognition thresholds.

In a standard multialternative forced-choice procedure (Blackwell, Note 1), one target stimulus and a number of blank stimuli are assigned randomly to different time periods or spatial locations. On each trial, the subject samples from each period, or location, and indicates which one contains the target. Since subjects are required to indicate where, rather than if, the target occurred, their willingness to report a perception is controlled. In addition, the target location for each trial is determined by a random process that makes each of the possible response alternatives equally likely to be correct on any given trial. Therefore, the disproportionate selection of a particular alternative that may result from a subject's response preference is no more likely to yield correct responses than is unbiased guessing.

Collings (1974) proposed a taste-recognition-threshold procedure based on the four-alternative version of the forced-choice procedure. In her procedure, stimuli representing each of the four taste qualities are randomly assigned to successive trials. On each trial, the subject tastes one sample and is forced to choose between four response alternatives—the four taste-quality names. Forcing the subject to respond with a quality name on each trial controls for variation in the willingness of subjects to report the perception of a taste quality. However, no mechanism controls for the preference of

I thank Stacey J. Arensen for testing the Study B subjects and Robert O Wolf for testing some of the Study A subjects. I greatly appreciate the critical comments of Beverly J. Coward on successive drafts of the manuscript. Requests for reprints should be sent to James M. Weiffenbach, National Institute of Dental Research, Building 10, Room 1-A-05, National Institutes of Health, Bethesda, Maryland 20205.

subjects for particular taste-quality names. Recall that the standard procedure controls for response preference by calculating a threshold from trials for which the various response alternatives were equally likely to be correct. The Collings procedure lacks this control mechanism because the target stimuli cannot be assigned arbitrarily to response alternatives. When the stimulus for sweet is presented, only the response "sweet" can be correct. For any given target, only one response is ever correct. Consequently, thresholds for each of the four qualities must be based on responses to a series of trials in which a particular response is invariably correct. This implies that a subject's preference for any particular response will artificially reduce the threshold value for the corresponding stimulus.

Since a subject's preference for one response over another can distort the threshold, it is of interest to know if preferential responding actually occurs. Presumably, response preference affects response not when the stimulus is clearly perceived but only when subjects lack sufficient stimulus information to respond correctly. Thus, in evaluating response preferences, attention has been focused on errors. Collings (1974) evaluated the distribution of erroneous responses (guesses) with a post hoc chi-square test and found no significant departure from chance expectation. Subsequently, similar tests failed to demonstrate differential guessing between comparison groups (McBurney & Moskat, 1975) or conditions (Collings, Lindberg, & McBurney, 1976). In contrast, Settle, Quinn, and Kare (Note 2) cited evidence suggestive of biased guessing. They observed that sweet and bitter thresholds correlated negatively with the ratio between the number of times each of these quality names was elicited as an incorrect response and the total number of incorrect responses. However, they did not report any analysis that would allow them to determine if sweet and bitter responses occurred disproportionately.

The possibility that preferential guessing may occur raises two issues. First, what are the characteristics of disproportionate selection of response alternatives? Second, does this behavior, when it occurs, invalidate the obtained thresholds?

The first issue may be reduced to two specific questions. Do the incorrect responses reflect individual differences in willingness to use the various responses or a uniform pattern of guessing that characterizes most subjects during some or all stages of testing? Is the pattern of guessing influenced by procedural variables such as the provision of information to the subjects concerning the correctness of their responses?

The issue of whether, or to what degree, valid threshold measurement is compromised by the subjects' response preferences cannot be resolved em-

pirically because no bias-free measurement procedure is available to generate comparison thresholds. However, an examination of the mechanisms by which response preferences affect measured thresholds may help investigators to resolve the practical question of whether to attempt to determine recognition threshold with the Collings procedure.

METHOD

Subjects

College students serving as normal volunteers at the National Institutes of Health Clinical Center acted as the subjects. At the time of testing, none were participating in any studies involving drugs, hormones, or special diets. They were instructed to refrain from smoking, eating, or drinking anything except water for 1 h before being tested.

The 19 subjects for Study A consisted of 8 male and 3 female subjects who had been examined and found to be free of dental caries, along with 5 male and 3 female subjects who achieved a DMF (decayed, missing, or filled) count of 10 or greater on similar examination. All subjects had previously been tested for taste-detection thresholds for sucrose, sodium chloride, citric acid, and quinine sulfate by a two-alternative forced-choice procedure.

The 12 subjects for Study B were equally divided between males and females, but no special selection was made with respect to caries status.

Stimuli

Four series of stimulus fluids were prepared, one each for sweet (sucrose), salty (sodium chloride), sour (citric acid), and bitter (monohydrate quinine sulfate). Reagent-grade chemicals and double-distilled water were used throughout. Successive solutions in each stimulus series differed by a quarter log unit of molar concentration. The extreme values for the series were as follows: 5.6×10^{-1} and 1.0×10^{-5} M for sucrose and sodium chloride, 1.0×10^{-2} and 1.0×10^{-6} M for citric acid, and 1.0×10^{-4} and 3.0×10^{-9} M for quinine sulfate. Fluids were stored in glass-stoppered reagent bottles under refrigeration (4°C) and brought to room temperature before use.

Procedures

Testing was performed after the manner of Collings (1974). Test stimuli representing each of the four basic qualities were presented one at a time in a predetermined random order without any water or blank stimuli. The subjects tasted each sample and then spit it out and assigned a taste-quality name to it. Only the four conventional taste-quality names were accepted as responses. The concentration of each test stimulus was determined by the subject's last response to a stimulus of the same quality. If the previous response was correct, the concentration was decreased. If it was incorrect, the concentration was increased. Thus, for each stimulus quality, the concentration of successive samples consistently increased or decreased until a change in the correctness of the response reversed the direction of stimulus-concentration change. Testing continued until seven or more such reversals had been obtained for each quality.

The subjects were required to rinse their mouths with distilled water before tasting each sample. They were provided with a 150-ml clear plastic tumbler and with bottles of distilled water from which to refill it.

In Study A, all the subjects were tested with 10-ml samples presented in 30-ml plastic medicine cups, and none were told whether their responses were correct or incorrect. Two different rules were used to terminate the testing session. For 9 of the subjects, stimuli representing each quality were presented until the criterion of seven stimulus-concentration reversals had been

achieved for all qualities. For the other 10 subjects, stimulation was terminated independently for each quality as soon as the seven-reversal criterion for that quality was reached.

In Study B, equal numbers of males and females were assigned randomly to treatment conditions that differed with respect to both the information the subjects were given about the correctness of their responses and the volume of stimulus fluid supplied for each test. All Study B subjects were told after each response whether that response had been correct or incorrect. Half the subjects were also told what quality name would have been correct. One third of the subjects in each group sipped samples of approximately 10 ml from plastic medicine cups. Another third had samples of .05 ml pipetted onto their tongues, and the remaining subjects received .005-ml samples. The subjects were tested using the same stimulation technique and the same conditions of informational feedback during a session on the following day. Second sessions were carried out at the same time of day, except that one subject who had been tested at 11:30 a.m. was retested at 2:00 p.m. and one subject who had been tested at 2:00 p.m. was retested at 11:00 a.m.

RESULTS

The frequency with which individual subjects gave each of the four taste-quality names as an incorrect response was tabulated. Overall, sour was the most frequent incorrect response and salty the least frequent error. This pattern was also observed when Study A and Study B were considered separately and when Study B subjects were subdivided into groups based on day of testing or on amount of informational feedback provided. To test for differences between quality names, the frequency scores of each subject were converted to the ranks 1 through 4. The average rank was highest for sour and lowest for salty, overall and for each of the subgroups considered above. The overall difference between quality names was significant by Friedman two-way analysis of variance [$\chi^2(3) = 12.19$, $p < .01$]. In separate tests for Studies A and B, the differences observed in Study B were significant [$\chi^2(3) = 10.43$, $p < .02$], whereas those in Study A were not [$\chi^2(3) = 3.96$, n.s.].

The frequency with which the four taste-quality names occurred as incorrect responses was tested against the expectation of equal occurrence by chi-square analyses carried out separately for each subject. In Study A, 5 of the 19 subjects (i.e., more than one-quarter of the sample) displayed departures from chance expectation significant beyond the .05 level. Groups formed on the basis of rules for terminating the testing session each contained individuals with significant chi-square values (2 of 9 and 3 of 10). In Study B, only 1 of the 12 subjects displayed significantly disproportionate responding on the first day of testing. However, on the second day of testing, this subject and four additional subjects (i.e., more than one-third of the sample) showed disproportionate responding significant beyond the .05 level. When data from Day 1 and Day 2 of Study B are combined, 5 subjects who showed dis-

proportionate responding on Day 2 are joined by 4 others. Thus, 9 of the 12 Study B subjects achieved chi-squares significant at the .05 level. These subjects were found in each sex group (4/6 males, 5/6 females), each stimulus-volume group (2/4 cups, 4/4 large drops, 3/4 small drops), and each information-feedback group (5/6 full feedback, 4/6 correctness only). In five cases, the chi-square was significant beyond the .01 level.

Disproportionate occurrence of taste-quality names as incorrect responses characterizes both the individual and group data. However, the pattern of disproportionate occurrence emerging from these two types of data differ. Whereas the group data indicate that sour is the most frequent incorrect response, only half of the 14 individual subjects showing significantly disproportionate responding have sour as their most frequent incorrect response. Similarly, salt, which is the least frequent incorrect response overall, is the least frequent error for only 5 of these 14 subjects. Salt was the least and sour the most frequent response for only 3 of the subjects with significant disproportionate responding. Substantial numbers of subjects exhibiting disproportionate responding show patterns that are different from that defined by the group average.

DISCUSSION

Incorrect responses elicited by taste stimuli during measurement of the taste-recognition threshold by the Collings (1974) procedure are assumed to occur as a result of the subject's inability to recognize the stimulus and to be a product of guessing. If such guessing were random, the four possible response alternatives would be equally likely to occur as errors. However, the distribution of errors observed in the present study departs significantly from this expectation. Thus, the characteristics of biased guessing and its possible effect on taste-recognition thresholds must be examined.

Disproportionate occurrence of taste-quality names as incorrect responses reflects the subjects' biased guessing and is a persistent feature of my experience with the Collings procedure. At least some individuals demonstrated significant nonrandom responding in their initial testing session. On subsequent retest, the number of subjects responding in this manner increased. None of the different procedural variations employed entirely eliminated biased use of response alternatives.

The effect of biased guessing is illustrated by a simplified worst-case situation. Imagine that you are measuring the threshold for only two qualities, say, bitter and sweet, and are testing a subject who prefers the response "bitter." Given the sweet stimulus, the subject responds "sweet" whenever that taste experience is clearly elicited. As correct re-

sponding continues, the concentration of the sweet stimulus is reduced to the point at which guessing is called for. Under these circumstances, this subject, unlike the unbiased subject, is much more likely to respond "bitter" than "sweet." Since this is an incorrect response, the downward progression of stimulus concentrations is terminated prematurely and the sweet threshold is elevated above what it would be in the absence of bias. When the bitter stimulus is present and recognized, the subject responds "bitter." Since this response is correct, the concentration of the stimulus is lowered. At some point the concentration is too low to elicit an experience of bitter and the subject is forced to guess. Due to bias, the subject is more likely to respond "bitter" than "sweet." Guesses in the presence of the bitter stimulus are more likely than chance to be correct, and, in the worst case, the threshold is depressed without limit.

Having looked at the simplest case, now consider what happens when a subject biased toward bitter is tested by the Collings procedure. The subject correctly identifies all stimuli regardless of quality when they elicit a clear, subjective taste-quality experience. However, when the experience elicited is ambiguous, the subject's bias makes the response "bitter" more likely than any other. Adding the two additional response categories has changed little. Thresholds for salty, sour, and sweet may be slightly elevated, but, in the worst case, the threshold for bitter will again be depressed without limit.

Insofar as the distribution of the subject's "guesses" departs from chance, threshold measurement with the Collings procedure approaches this worst-case situation. Although incorrect responses are not a perfect index of a subject's "guesses," my data strongly support the contention that subjects, faced

with insufficient sensory information, do not guess randomly. Thus, the failure of the Collings procedure to compensate for response bias becomes critical and is a strong counterindication for the use of this method.

REFERENCE NOTES

1. Blackwell, H. R. *Psychophysical thresholds: Experimental studies of methods of measurement* (Bulletin of the Engineering Research Institute, No. 36). Ann Arbor: University of Michigan, 1953.
2. Settle, R. G., Quinn, M. R., & Kare, M. R. *Report for the August 4, 1976, Meeting on Gustatory Evaluation of Cancer Patients* (Contract No. Z-CP-65791). Philadelphia, Pa: Monell Chemical Senses Center, 1976.

REFERENCES

- COLLINGS, V. Human taste response as a function of locus of stimulation on the tongue and soft palate. *Perception & Psychophysics*, 1974, **16**, 169-174.
- COLLINGS, V. B., LINDBERG, L., & MCBURNEY, D. H. Spatial interactions of taste stimuli on the human tongue. *Perception & Psychophysics*, 1976, **19**, 69-71.
- HARRIS, H., & KALMUS, H. The measurement of taste sensitivity to phenylthiourea (PTC). *Annals of Eugenics*, 1949, **15**, 24-51.
- JONES, F. N. A forced-choice method of limits. *American Journal of Psychology*, 1956, **69**, 673-674.
- MCBURNEY, D. H., & MOSKAT, L. J. Taste thresholds in college smokers and non-smokers. *Perception & Psychophysics* 1975, **18**, 71-73.
- WOTMAN, S., MANDEL, I. D., THOMPSON, R. H., JR., & LAVAGH, J. H. Salivary electrolytes and salt taste thresholds in hypertension. *Journal of Chronic Diseases*, 1967, **20**, 833-840.
- YASAKI, T., MIYASHITA, N., AHIKO, R., HIRANO, Y., KAMATA, M., & IZUKA, Y. Study on sucrose taste thresholds in children and adults. *Japanese Journal of Dental Health*, 1976, **26**, 20-25.
- ZENGO, A. N., & MANDEL, I. D. Sucrose tasting and dental caries in man. *Archives of Oral Biology*, 1972, **17**, 605-607.

(Manuscript received February 18, 1982;
revision accepted for publication December 13, 1982.)