

# The roles of practice and correction in the categorization of sour and bitter taste qualities

WAYNE K. McAULIFFE and HERBERT L. MEISELMAN

*Pioneering Research Laboratory, U.S. Army Natick Laboratories, Natick, Massachusetts 01760*

Ten Ss were selected from a group of 40 screened to determine misnaming of stimuli usually called sour (hydrochloric and citric acid) and bitter (quinine sulfate). The 10 Ss were then given 16 trials in which they were asked to match the stimuli to standards, with or without verbal feedback for each stimulus. This was followed by 16 trials without feedback for all Ss and, several days later, by a 16-trial posttest. Final improved performance of the Ss was not affected by having received the feedback. It was concluded that experience with the stimuli leads to a reduction in the degree of misnaming, implying a nonphysiological basis for sour-bitter confusions.

Use of the traditional taste quality labels of sour, salty, bitter, and sweet is marked by relatively high inaccuracies in assignment of sour and bitter labels to taste stimuli. This highly variable assignment of taste quality names ("mislabeled") for sour and bitter is in contrast to observations for stimuli usually assigned the quality labels of sweet or salty (Amerine, Pangborn, & Roessler, 1965, p. 108).

During the development of a gustatory screening test, Meiselman and Dzendolet (1967) observed that, with 10-ml sips of 20 mM HCl and 2 mM KCl or with 20 mM HCl and 0.0008 mM QSO<sub>4</sub>, 28% of 60 male Ss and 10% of 60 female Ss mislabeled the sour and bitter qualities. Meiselman and Dzendolet attempted to change this misnaming with a correction procedure in which the S was informed after his first stimulus how the average S would respond, but he was encouraged to report his own opinion. No significant change in quality responding was observed, leading the authors to suggest the possibility of a physiological basis for the observed misnaming.

Sato (1971) has suggested, based on neurophysiological recording from chorda tympani fibers in rat and hamster, that three groups of fibers exist, one sensitive to hydrochloric acid, quinine, and cooling, another sensitive to sodium chloride, and a third to sucrose and warming. Although comparable neurophysiological data in humans do not exist, Sato suggests that the human sour-bitter misnaming phenomenon might reflect this underlying physiological grouping.

Robinson (1970) also has reported the sour-bitter misnaming using quality responses of 48 Ss to single drops of 2.4 mM citric acid and 0.013 mM QSO<sub>4</sub>. Twenty-two percent assigned the sour label to QSO<sub>4</sub>, and 25% assigned the bitter label to citric acid. Fifty-six Ss (including the above 48) indicated in a paper-and-pencil survey what they thought would be the predominant taste of a variety of foods. Robinson

suggests that those foods with predominantly sour tastes (lemons, cooking apples, lime juice, rhubarb) and bitter tastes (pale ale, strong black coffee) were the least consistently identified, although all foods appeared to include a profile of taste quality responses. For example, 33% identified the predominant taste of lemonade as bitter and 39% did the same for lime juice. Conversely, 14% identified strong black coffee as sour. Robinson reported that those individuals who had mislabeled the simple solutions mislabeled the complex foods.

Robinson (1970) argues that the physiological color blindness analogue suggested by Meiselman and Dzendolet (1967) is incorrect, and that the sour-bitter misnaming phenomenon is, in fact, due to a language or naming problem.

Most recently, Gregson and Baker (1973) used a rating scale for measurement of perceived intensity of the sourness and bitterness of near-threshold solutions of citric acid, quinine sulfate, and disodium-5'-guanylate, each prepared in a seven-step series. Gregson and Baker developed a psychophysical parameter which indicates the degree of sour-bitter confusion, varying from complete confusion to discrimination. Using this approach, they determined that 3 in 15 Ss showed complete sour-bitter confusion with quinine sulfate, 1 in 14 for citric acid, and 3 in 16 for disodium-5'-guanylate.

Aside from introducing the intensity scaling to the phenomenon of sour-bitter misnaming, Gregson and Baker have approached the issue from a different philosophical basis. They argue that the misnaming phenomenon is based in the aqualitative intensity perception which precedes quality-related intensity in taste. The sour-bitter misnaming occurs because Ss slip back to aqualitative intensity rating when a quality is confusing or secondary.

The present experiment was undertaken to determine whether Ss with a demonstrated sour-bitter

misnaming propensity could be trained to "correctly" name the solutions and whether this training might have a lasting effect.

**METHOD**

**Subjects**

Forty Ss were selected from the 550-person taste-testing panel at the U.S. Army Natick Laboratories and screened, leaving 16 who qualified for further testing. Of the 16 who qualified, 10 agreed to participate in the present study, 7 males and 3 females ranging in age from 24 to 56 years.

**Stimuli**

The solutions used in both the screening task and throughout the remainder of the experiment were 20 mM HCl and 24 mM citric acid to represent sour and 0.013 and 0.008 mM quinine sulfate (QSO<sub>4</sub>) to represent bitter. All solutions were made with reagent-grade chemicals and demineralized water. The solutions were chosen to allow comparison of the stimuli in the Robinson (1970) and Meiselman and Dzendolet (1967) studies.

**Procedure**

A screening task was used to find Ss who could not reliably assign the words "sour" and "bitter" to the acids and QSO<sub>4</sub>, respectively, when presented with 16 randomly ordered test solutions (four of each stimulus). A response of "no taste" was permitted. A criterion of six errors was used to define a sour-bitter confusion.

The 10 Ss who qualified for and agreed to further testing were divided into two groups of 5, matched according to screening task scores. Where unequal scores were matched, the S with the higher score was assigned to the experimental group. There was no significant difference between experimental and control group screening task scores ( $t = 0.305, df = 9$ ).

Both experimental and control groups were then tested in a 32-sample training session, 4 of each stimulus randomly ordered within 2 groups of 16. Both groups were also given four standards labeled "A," "B," "C," and "D," each letter representing one of these standard solutions. All Ss were allowed to sample the standards only once at the beginning of the session. The task was to label the test solutions (numbered 1 to 32) with the letter of the standard that tasted most similar, and to try and make as many correct responses as possible.

**Table 1**  
Flow Chart of the Experiment Showing Breakdown of the Three Sessions and 64 Trials

Session	Trial	N of Ss	Task
1 (Screening)	1-16	40	Assign sour and bitter labels to 16 stimuli.
2 (Feedback)	17-32	5	Assign labels of standards (A, B, C, D) to 16 stimuli with feedback on each stimulus.
	17-32 (No Feedback)	5	Assign labels of standards (A, B, C, D) to 16 stimuli without feedback on each stimulus.
	33-48	10	Assign labels of standards (A, B, C, D) to 16 stimuli without feedback on each stimulus.
3 (Posttest)	49-64	10	Assign labels of standards (A, B, C, D) to 16 stimuli without feedback on each stimulus.

*Note*—Ss were treated differentially only during Trials 17-32 of Session 2. Note that the taste was different in Session 1 from that in Sessions 2 and 3.

**Table 2**  
Percentage of Response for Four Compounds and Three Response Categories (Sour, Bitter, No Taste) from Initial Screening

	Sour	Bitter	No Taste
HCl	45.4	35.9	18.7
Citric Acid	67.2	29.7	3.1
Quinine Sulfate .008 mM	7.8	23.3	68.9
Quinine Sulfate 13 mM	7.8	18.8	73.4

*Note*—Each value represents the percentage of response for 40 Ss tested with four presentations of each of the four compounds listed.

The feedback group was given verbal feedback after each of the first 16 samples. They were told they were "correct" when they made a proper comparison or they were told "it should have been called \_\_\_\_\_" (the correct letter was supplied) when they made an incorrect comparison. For the last 16 samples, the Ss responded without feedback.

The nonfeedback group was allowed to classify all 32 samples without feedback.

All Ss were given a posttest of 16 additional samples from 5 to 14 days after the 32-sample training session. They tasted the standards only once, at the beginning of the session, and no feedback was given to either group.

A flow chart of the experiment is presented in Table 1. Throughout the screening task and experiment, the Ss were instructed to allow 30 sec between samples.

A sip-and-spit procedure was used to present the solutions. Responses were recorded by the Ss themselves during the screening task and by the E during the training and posttest.

**RESULTS**

In the original 40-person group (Table 2), 68.9% of the Ss responded "no taste" to the weaker quinine solution while 73.4% responded "no taste" to the stronger quinine solution; only 7.8% responded "sour" to either quinine solution; 45.4% of the Ss responded "sour" to HCl, while 67.2% responded "sour" to citric acid; 35.9% responded "bitter" to the HCl, and 29.6% responded "bitter" to the citric acid.

The combined feedback and nonfeedback data show a significant decrease in errors from the 16 samples of the screening task to the posttest ( $F = 7.065, df = 3/36, p < .01$ ).

When considered alone, the feedback group showed a significant decrease in errors over trials ( $F = 4.686, df = 3/16, p < .05$ ), but the nonfeedback group did not ( $F = 2.935, df = 3/16$ ). There was a significant difference between the screening task and the first 16 trials of the training session for the control group ( $t = 2.5457, df = 9, p < .05$ ) as well as between the screening task and the second 16 training trials ( $t = 2.5323, df = 9, p < .05$ ), but there was no significant difference between the training session and the posttest ( $t = 0.078, df = 9$ ) (see Fig. 1). The increase in the number of errors in the third block of trials was not statistically significant.

One of the Ss in the experimental group was unable to complete the posttest, so the mean of the remainder

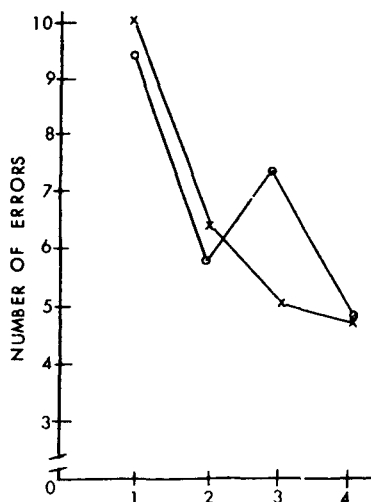


Fig. 1. Number of errors in assigning sour and bitter categories to solutions of acid and quinine as a function of trials (in blocks of 16). Those Ss given a correction procedure in the second block of trials are shown by crosses, and those in the noncorrection group are shown by circles. The first block of 16 trials represents screening; the second block, training with or without correction; the third block, testing after correction; and the fourth block, testing after an interval of 5-14 days.

of the group was added once to the group data to even the control and experimental condition sizes.

The present experiment produced different percentages of misnaming than did prior studies, although the reader should remain aware of sampling differences in these studies. Generally, there was a much smaller percentage (7.8%) of reports of sour to quinine sulfate than reported by Robinson (22%) or Meiselman and Dzendolet for potassium chloride (average of 17.3%). Gregson and Baker (1973) reported 20% total confusion for quinine sulfate. The percentage of reports of bitter to acid increased, but less markedly; the present study obtained 29% and 35% calling citric acid and hydrochloric acid, respectively, bitter. Meiselman and Dzendolet obtained 28.4%, on the average, for HCl, and Robinson reported 25% for citric acid, but Gregson and Baker reported only 7% total confusion for citric acid. The high percentage of reports of no taste for the quinine stimuli in the present study might have contributed to the low percentage of reports of sour.

The percentage of those who tasted anything (sour or bitter) who responded sour to quinine was at least 25%. A slightly higher quinine level would probably have yielded higher percentages of those reporting bitter and sour.

The basic finding of the training procedure is that experience with different compounds, rather than

instructions or feedback, produces a significant and lasting effect on taste quality categorization. Figure 1, which is divided into blocks of 16 trials for convenience, shows the dramatic decrease in errors when training commenced after initial screening. Although not at the level of statistical significance, the data suggested continued improvement in taste quality categorization after training ended.

In terms of the criteria used in this study to define misnaming, those Ss who misnamed and were chosen for study reached the criterion of correct naming (10 out of 16 correct) during the first block of training trials, and maintained improvement.

Meiselman and Dzendolet demonstrated that correction on the first trial produced no improvement in sour-bitter labeling. The present study demonstrated that the final level of performance was not different for feedback and nonfeedback groups. Both showed a significant and nontransitory reduction in errors, although only one had received a correction procedure. However, the correction procedure in the present study differed quantitatively from that used by Meiselman and Dzendolet; correction was given on every one of 16 trials. It appears that repeated opportunities to sample from compounds representing sour and bitter taste qualities, i.e., practice, accounts for the final performance of both groups. Perhaps practice and an extensive correction procedure are functionally the same thing. In the terminology of Gregson and Baker, exposure to appropriate stimuli overcomes confusion and improves performance above the aqualitative level.

## REFERENCES

- AMERINE, M. A., PANGBORN, R. M., & ROESSLER, E. B. *Principles of sensory evaluation of food*. New York and London: Academic Press, 1965.
- BARTOSHUK, L. M. Water taste in man. *Perception & Psychophysics*, 1968, **3**, 69-72.
- GREGSON, R. A. M., & BAKER, A. F. H. Sourness and bitterness: Confusions over sequences of taste judgments. *British Journal of Psychology*, 1973, **64**, 71-76.
- MEISELMAN, H. L., & DZENDOLET, E. Variability in gustatory quality identification. *Perception & Psychophysics*, 1967, **2**, 496-498.
- ROBINSON, J. O. The misuse of taste names by untrained observers. *British Journal of Psychology*, 1970, **6**, 375-378.
- SATO, M. Neural coding in taste as seen from recordings from peripheral receptors and nerves. In L. Beidler (Ed.), *Handbook of sensory physiology*. Vol. IV. *Chemical senses*. Part 2. Berlin: Springer-Verlag, 1971. Pp. 116-147.

(Received for publication December 17, 1973;  
revision accepted May 13, 1974.)