Chemical Senses and Flavour Volume 4 Number 4 1979

Confusion in the use of the taste adjectives 'sour' and 'bitter'

M. O'Mahony, M. Goldenberg, J. Stedmon* and J. Alford

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

(Received June 1979; revised September 1979; accepted October 1979)

Abstract. Subjects from Britain and America were tested on their skill at applying the taste adjectives 'sweet', 'sour', 'salty', and 'bitter' to clearly distinguishable solutions of sucrose, citric acid, NaCl and quinine sulphate, respectively. The main error that occurred was calling citric acid 'bitter' while the tendency to call quinine sulphate 'sour' was not so common; this is the well known sour-bitter confusion. A sour-salty confusion was also noted as well as a tendency to call citric acid 'sweet'. All these confusions were rectified by mere definition using standards. Skill at applying taste adjectives was not always found to be consistent over time. More errors occurred at lower solution concentrations, even though stimuli were clearly distinguishable; indistinguishability of stimuli may account for some confusions in other studies.

1. Introduction

Unlike vision, taste has a sparse descriptive language in Western culture; even preschool children are far better at color naming than taste naming (O'Mahony *et al.*, 1978), probably due to practice given by their parents. Because of this, sensory analysts have developed a range of techniques for the description of food flavours which involve the definition of descriptive terms using physical standards (e.g. Caul, 1957; Caul *et al.*, 1958; Stone and Sidel, 1974; Williams, 1975).

Psychophysicists, however, have generally made the implicit assumption in their methods that there is a concensus of agreement regarding the use of the taste adjectives, 'sweet', 'sour', 'salty', and 'bitter' (tasteless also) because they rarely give subjects training sessions to match these terms to standards. Some research is more concerned with how subjects apply these terms without prior definition by matching to standards but usually this is not the case. That subjects will apply taste adjectives in the same way as the experimenter is an assumption.

A commonly noted case of the wrong application of adjectives in English is the so-called sour-bitter confusion (Myers, 1904; Pangborn, 1961; Amerine *et al.*, 1965); other languages have their own taste confusions (Chamberlain, 1903; Myers, 1904). The English sour-bitter confusion has been studied by workers in Britain, New Zealand and U.S.A. (Meiselman and Dzendolet, 1967; Robinson, 1970; Gregson and Baker, 1973; McAuliffe and Meiselman, 1974).

Meiselman and Dzendolet (1967) elicited from 80 subjects taste descriptions, limited to 'sweet', 'sour', 'salty', 'bitter' and 'no taste', for 10 ml sips of 15 mM sucrose, 50 mM NaCl, 2 mM HCl and 20 mM KCl. Using a criterion of at least 4 correct out of 7 presentations, only 30% of the subjects exceeded the criterion for all stimuli; all types of confusions were made but sour-bitter confusions were most common ('sour' called 'bitter' more than vice versa) occurring for 21.25% of the subjects. In an additional experiment, 20 subjects described in the same way 25mM sucrose, 40 mM NaCl, 2mM HCl and 8μM quinine sulphate. A further 20 subjects performed this experiment except that after one presentation of each of the solutions, they were told how the average subject would have responded, although it was emphasised that subjects should report their own opinions. Using a criterion of 4 correct out of 6 presentations, 20% of the first group made correct identifications for all stimuli and 35% of the latter group. The latter group thus performed slightly better but not totally correctly. All types of taste confusions were made but sour-bitter confusions were most common (again, 'sour' called 'bitter' more than vice versa), occurring for 10% of the prior group and surprisingly 35% of the latter 'informed' group. The authors suggested that the lack of success of the correction procedure may indicate the possibility of a physiological basis for the misnaming, analogous to abnormal colour vision. It should be noted, however, that the 'correction procedure' did not train subjects in the definition of taste adjectives; in fact, it emphasized that subjects should use their own descriptions. Further, a high frequency of 'no taste' descriptions in both experiments suggests that some of the stimuli may have been of too low a concentration to be distinguished in this setting.

Robinson (1970) elicited from 48 subjects taste descriptions ('sweet', 'sour', 'salty', 'bitter', their combinations or 'no taste') for single drops of 24 mM citric acid and 130 μM quinine sulphate. All possible taste confusions were made, although only the quinine was called tasteless. The sour-bitter confusion was the most common; citric acid was called bitter for 25% of the judgements and quinine sour for 22%. These subjects were chosen from a pool of 56 who were required to name which of the four 'primary' (sweet, sour, salty, bitter) tastes was the predominant taste for a list of foods. Again, subjects made 'confusions' here, while there was a trend for those who wrongly named solutions to wrongly name foods. He concluded that the mistakes were more like lack of skill in naming rather than physiological effects analogous to colour blindness.

Gregson and Baker (1973) required subjects to use a ratio range category scale to estimate the perceived sourness and bitterness of 5 ml samples of ascending series of concentrations of citric acid, quinine sulphate and disodium-5′-guanylate (D-5′-G); each series had seven concentrations (including water) doubling in value ranging for the three stimuli 0.37-23.79 mM, 2.79-178.81 μM, 0.58-36.84 mM, respectively.

They analysed their results by fitting a psychophysical model of qualitative confusions and developing a parameter, θ, which indicated the degree of confusion between two taste qualities. They found complete sour-bitter confusions for three in 15 subjects for quinine and one in 14 for citric acid. For some subjects there was a tendency to rate quinine as sour and citric acid as bitter while for others there was discrimination. They felt that there was no justification for treating the confusion as purely psychological or physiological and that the important point posed by their results was that aqualitative (total without regard to specific taste quality) taste intensity estimation may precede estimation for specific tastes. They stated that confusions would seem to occur when subjects faced with rating a quality, which was secondary or which they were not sure was present, reverted to making aqualitative estimates of total intensity so giving inappropriately high bitter ratings for citric acid and sour ratings for quinine. Their explanation would seem specific to confusion as defined more complexly by their experiment rather than the application of inappropriate adjectives. It is also possible that the perceived demand characteristics of the subjects' task were such that the presence of 'inappropriate' tastes were suggested by the instruction to make intensity ratings for 'inappropriate' taste qualities.

McAuliffe and Meiselman (1974) obtained 10 subjects with sour-bitter confusions, from a group of 40, by a screening task requiring subjects to assign the descriptions 'sour', 'bitter' or 'no taste' to four presentations each of 20 mM HCl, 24 mM citric acid, 13 μM and 8 μM quinine sulphate. A criterion of six errors was used to define a sour-bitter confusion. The subjects were divided into two equal groups and required to match eight each of the four solutions (32 samples) to four standards tasted before the sessions; one group was given no feedback, the other was given feedback for the first 16 samples. Both groups were tested a few days later using the matching procedure without feedback.

In the original screening task 35.9% of subjects responded 'bitter' to HCl and 29.6% to citric acid while only 7.8% responded 'sour' to either quinine solution. A substantial number of 'no taste' responses also suggested that some of the confusions may have been due to the indistinguishability of the stimuli in this setting. They found that errors tended to decrease throughout and sessions for both feedback and non feedback groups and concluded that experience with the stimuli rather than instructions or feedback improved the subjects' matching.

Although these four studies were concerned with the sour-bitter confusion their contexts were different. The first two (Meiselman and Dzendolet, 1967 Robinson, 1970) were concerned with subjects' skill at applying appropriate adjectives to taste sensations. However, the latter two studies did not address this. Gregson and Baker (1973) investigated intensity scaling along inappropriate dimensions while McAuliffe and Meiselman (1974) examined subjects' skills in matching taste sensations. This study, like the former two, concerns itself with the practical methodological problem of how simple definition of taste adjectives by presentation of appropriate standards, corrects their misapplication. Further, it notes the effect of stimulus concentration on error rate and, as a control, it also investigates whether stimuli used by prior investigators were distinguishable.

2. Method

All stimuli used in the study were, unless otherwise stated, reagent grade chemicals dissolved in water purified by a Milli-RO 4/5 Filtration and Reverse Osmosis System, in series with a Milli-Q purification system: ion exchange and activated charcoal (Milli-pore Corp., Bedford, Mass.). Solutions were stored in previously steamed, odourless polythene bottles and presented, unless otherwise stated, at room temperature (18-26°C) in 1 oz unwaxed paper cups (Portion Control Cups, American Convenience Products, Milwaukee, Wisconsin) coded with 3-digit random numbers.

All 882 subjects were unpaid students, staff and friends of University of California, Davis (U.C.D.), U.S.A. except for those 100 subjects in Experiment 5 who were undergraduates at Bristol University, England. No subjects participated in more than one experiment except for those in Experiment 4 who had previously participated in Experiment 1. Subjects were told to taste stimuli carefully, allowing them to reach the back as well as the front of the mouth.

2.1 Experiment 1

45 subjects (17M, 28F, 20-27 yrs), Food Science and Technology (FST) students and a further 100 subjects (47M, 53F, 17-31 yrs) who were not Food Science students (non-FST), were seated in sensory booths; they tasted (sip and expectorate) 15 ml samples of 300, 350, 400 mM NaCl, 300 mM sucrose, 20, 25 mM citric acid and 200 μM quinine sulphate (QSO_4) solutions, in random order. More than one of some of these stimuli were included making 12 samples in all. The numbers of each are given for this and succeeding experiments in Table I (column 1a for this study). Subjects responded by writing the corresponding 3-digit number in the appropriate column of a response sheet labelled 'sweet', 'sour', 'salty', 'bitter', 'peppermint', 'tasteless'. The last two were included to minimise expectancy effects while 'tasteless' also served to check that tastes were perceived. The odd numbers of samples were used to

Table I. Number of each stimulus concentration used for Experiments 1-6.

	Experiment																		
	1				2						3	4				5		6	
Stimuli (mM)	1a	1b	1c	1d	2a	2b	2c	2d	2e	2f	3a	4a	4b	4c	4d	5a	5b	6a	6b
NaCl: 400	1	1	1	1								1						1	1
350	2	1	1	1								1							
300	2	1	2	1	1	1		1		1		1	1			1	1		
250					1	1	1	1	1	1			1	1		1	1		
200					1	1	1	1	1	1			1	1		1	1		
150					1		1	1	1	1			1	1		1	1		
100					1		1						1	1		1	1		
75															1				
50											1				2				
Sucrose: 300	1	3	2	3								3						1	1
200																1	1		
100					1	3	2	2	3	2			1	2	1				
15											1				1				
Citric 25	1	1	1	1								1						2	1
Acid: 24											1				2				
20	2	2	2	2	1	1	1	1	1	1		2	1	1		1	1		
15					1	1	1	1	1	1			1	1		1	1		
10					1	1	1	1	1	1			1	1		1	1		
8															1				
Quinine 0.200	3	3	3	3								3						2	1
Sulfate: 0.178											1				1				
0.130											1				1				
0.050					1	1	1	1	1	1			1	1		1	1		
0.045					1	1	1	1	1	1			1	1		1	2		
0.040					1	1	1	1	1	1	1		1	1		1			
0.013											1				1				
0.008											1				1				
Total Stimuli:	12	12	12	12	12	12	12	12	12	12	8	12	12	12	12	12	12	6	4

minimise expectancy effects and the order of categories on the response sheet was randomised (many orders) to minimise bias.

Verbal as well as written instructions were given to each subject and experimentation was not started until subjects understood the procedure completely. Between tastings, subjects rinsed thoroughly (at least one rinse, no limit to number) in their own time, to minimise effects of adaptation to residuals (O'Mahony, 1972a, b, 1973; O'Mahony and Dunn, 1974; O'Mahony and Godman, 1974; O'Mahony and Wingate, 1974). Prior control experiments involving sorting, had determined that using this procedure, the sweet, sour, salty and bitter stimuli could be clearly distinguished as separate categories without error. Further controls indicated that judges (N = 10) practised and familiar with the taste categories agreed that the stimuli were clearly typical.

Subjects who agreed to return were then recalled 3-4 weeks later and required to repeat the experiment; the only change was that the number of stimuli for each concentration had changed, so as avoid any bias due to subjects' expectancies (see Table I, column 1b).

Those subjects who had been consistently incorrect in the initial and recall trials 3-4 weeks later were given a learning trial immediately after their recall trial. 15 ml samples of the four stimuli (400 mM NaCl, 200 mM Sucrose, 25 mM citric acid, 200 μM QSO_4) were presented in random order in 1 oz paper cups for tasting (sipping, expectoration, interstimulus tapwater rinsing) along with their respective adjectival descriptions: 'salty', 'sweet', 'sour', 'bitter', respectively. Subjects tasted the stimuli with feedback as often as they wished until they felt confident of their discriminative and descriptive abilities. Following the learning trial subjects returned to the booth for a further test session following exactly the same procedure as the first part of Experiment 1 except that there were different numbers of each stimuli (see Table I, column 1c). Those subjects who did not get a perfect score repeated the learning trial and a further test trial with the numbers of each stimulus distributed according to Table I, column 1d.

2.2 Experiment 2

This experiment was essentially a repeat of Experiment 1, except that stimuli of generally lower concentrations were used, viz: 100, 150, 200, 250, 300 mM NaCl, 100 mM sucrose, 10, 15, 20 mM citric acid, 40, 45, 50 μM QSO_4. 100 non-FST subjects (42M, 58F, 17-45 yrs) were tested for the first part of the experiment of whom approx. half agreed to return to be given learning trials if required. As a control, subjects were required to indicate any stimuli which had the same taste quality; all subjects indicated that they could distinguish the four 'primary' category stimuli. This control procedure was simpler and more efficient than a separate control experiment used in Experiment 1.

The distribution of stimuli used in the initial part of the experiment is given in Table I, column 2a. For retesting 3-4 weeks later the distribution is given in Table I column 2b for those who were initially correct and column 2c for those initially incorrect. After the first, second and third learning trials, the distributions were as shown in columns 2d, 2e, 2f; the stimuli used for the learning trials were 300 mM

NaCl, 100 mM sucrose, 20 mM citric acid, 50 μM QSO_4.

2.3 Experiment 3

50 non-FST subjects (16M, 34F, 18-35 yrs) tasted (sip and expectorate) even lower concentration stimuli than those used in Experiment 2: 50 mM NaCl, 15 mM sucrose, 24 mM citric acid, 8, 13, 40, 130, 178 μM QSO_4 distributed according to Table I, column 3. They were required to group the stimuli according to qualitatively different taste groups regardless of concentration. They could use as many groups as they wished.

2.4 Experiment 4

This experiment was a within subjects comparison of performance using the three concentrations used in Experiments 1, 2 and 3. 27 FST subjects (8M, 19F, 20-27 yrs) who had demonstrated initial errorless application of adjectivees when tested as in Experiment 1, were recalled 5 weeks later and tested as in the recall session for Experiments 1 and 2. Stimuli used for Experiment 1 (Table I, column 4a) were tasted followed immediately by a session using the lower concentration stimuli of Experiment 2 (Table I, column 4b). One week later the subjects returned to taste the concentrations used in Experiment 2 (Table I, column 4c) followed immediately by the still lower concentrations in Experiment 3 (Table I, column 4d). No feedback was given.

2.5 Experiment 5

This was essentially a near repetition of Experiment 2 except that the study was carried out using subjects who spoke English as spoken in England rather than in the U.S.A.

100 non-FST subjects (46M, 54F, 18-28 yrs), undergraduate students at Bristol University, England, tasted (sip and expectorate) solutions of NaCl, sucrose, citric acid and QSO_4 with concentrations identical to those for Experiment 2, except that a higher sucrose concentration was used (Table I, column 5a).

The other details of the experiment were as for Experiment 2 except that subjects were not placed in taste booths, were given verbal instructions only, but did respond in writing as for Experiment 2. 10 ml samples of stimulus were presented in 10 ml pyrex beakers, there being four random orders of presentation. Approx. 50 ml interstimulus tapwater mouthrinses were taken between stimuli. NaCl and sucrose were Analar grade and citric acid and QSO_4 were reagent grade chemicals dissolved in twice distilled water.

For the learning trials, the procedure was as for Experiment 2, except that reference standards were only presented for 'sour' (15 mM citric acid) and 'bitter' (45 μM QSO_4) and were 10 ml samples in clearly labelled 50 ml beakers. 10 subjects (6M, 4F, 19-22 yrs) were recalled 2-3 weeks later; they had misnamed sour and bitter stimuli to a criterion of at least 3/6 mistakes. Incidentally none had made errors for NaCl and two had misnamed sucrose (peppermint, tasteless). During retesting, the stimuli were as for the initial testing except that the 40 μM QSO_4 was replaced by 45 μM QSO_4 (Table I, column 5b).

A further 10 subjects (4M, 6F, 18-21 yrs), who made at least four miscategorisations in the initial testing but who had not been recalled for the learning trials were recalled 2-3 weeks later for re-testing a set of stimuli identical to those used in the initial testing. Immediately after, they were tested using higher concentration stimuli as used in Experiment 1 (Table I, Column 1a), without any learning trials. For both sets of stimuli the procedure was used as for the initial tasting.

2.6 Experiment 6

This experiment was a variation of Experiment 1. Subjects were first year students (age 18-20) who were surveyed in their rooms in halls of residence (dorms). 297 subjects were given, in random order, samples of 400 mM NaCl, 300 mM sucrose, 25 mM citric acid, 200 μM QSO_4, six samples distributed as in Table I, column 6a. The stimulus (approx. 5 ml) was squirted into the subject's mouth from a polythene squeeze bottle. The subject responded by stating whether it was 'sweet', 'sour', 'salty', 'bitter' and whether any of the samples tasted the same; those who reported that any stimuli tasted the same would have been dropped from the study, but none did. Those who made errors were retested immediately in the same way, using one of each stimulus (Table I, column 6b), to check that they were consistently wrong. Those subjects who consistently made errors were immediately given learning trials. They tasted the four stimuli ad lib and were told, as often as required, the 'correct' names for each. They were then tested again and, if necessary, given more learning trials until they were correct. For all these trials one of each stimulus was used (Table I, column 6b). Those subjects who had needed learning trials were tested again 3-4 weeks later using the same four stimuli, to see whether they had retained their taste naming skill.

2.7 Experiment 7

This experiment was a modification of Meiselman and Dzendolet's (1967) first experiment. 22 undergraduate students (13M, 9F, 18-25 yrs) tasted 2 mM HCl, 50 mM NaCl, 20 mM KCl and 15 mM sucrose (reagent grade chemicals in purified water equivalent to once distilled), in random order. As in Meiselman and Dzendolet's study the subject, after being seated at a sink and blindfolded, sipped 10 ml of solution from a 50 ml beaker, held it in his mouth approx. 3 secs and expectorated in the sink. Before tasting the next stimulus, he reported the taste quality, his responses being limited to 'sweet', 'sour', 'salty', 'bitter' or 'no taste'. No interstimulus rinses were taken although subjects expectorated as often as they wished. Time between presentations ranged 90-120 secs. The procedure was modified however in that each stimulus was only presented once and that besides reporting the taste quality subjects were required to state which stimuli, if any, tasted the same.

2.8 Experiment 8

This experiment was a modification of Meiselman and Dzendolet's (1967) second experiment. 27 subjects (7M, 20F, 18-24 yrs) tasted 2 mM HCl, 40 mM NaCl, 25 mM sucrose and 8 μM QSO_4, reagent grade chemicals in purified water equivalent to once distilled. 13 subjects tasted the stimuli according to Experiment 7 while the remaining 14 were tested according to Meiselman and Dzendolet's correction procedure; after tasting but before responding, subjects were told how the average subject would have responded but it was emphasised that he should report his own opinion. As in Experiment 7, subjects tasted each stimulus only once and were required to state which stimuli, if any, tasted the same.

2.9 Experiment 9

This experiment was a modification of Robinson's (1970). 32 subjects (8M, 24F, 17-21 yrs) tasted 24 mM citric acid followed by 130 μM QSO_4 by having a single drop of solution placed on the midline of the dorsal surface of the tongue approx. 2 cm from the top using a dropper; this position was adopted in the absence of details from Robinson. A 2 min interstimulus interval (no rinses) was used. Subjects were asked to identify the tastes of the solutions by choosing one or more of the descriptions 'sweet', 'sour', 'salty', 'bitter'; they could also respond 'no taste'. As in Robinson's experiment subjects first completed a food taste naming questionnaire. This was identical to that used by Robinson except the food 'lemonade' was replaced by '7-Up', to allow for language differences between Britain and U.S.A. The tasting procedure was modified in that subjects reported not only the taste quality of the two stimuli but also whether they tasted the same or different.

2.10 Experiment 10

This experiment was related to Gregson and Baker's (1973). 30 subjects (20M, 10F, 18-26 yrs), undergraduates naive to taste pshychophysical experimentation, were given a series of pair comparisons in which they were required to determine which of the pair was 'sour'; beforehand it was determined that subjects knew the meaning of 'sour'. The pairs of stimuli were comparable in concentration to the weakest solutions used by Gregson and Baker in their solution series. They were: 0.37, 0.74, 1.49, 2.97 mM citric acid paired with 2.79, 5.59, 11.18, 22.35 μM QSO_4, respectively.

5 ml of stimulus was presented to the subject in a 50 ml pyrex beaker. He took it all in his mouth, held it there for 5 sec, expectorated and tasted the comparison solution no sooner than after 30 secs. Subjects' times were strictly paced; stimuli were held at 25°C in a water bath and tasted immediately after removal from the bath.

2.11 Experiment 11

This experiment was related to McAuliffe and Meiselman's (1974). 29 other subjects (10M, 19F, 18-21 yrs) were required to taste, in random order, 20 mM HCl, 24 mM citric acid, 13 μM and 8 μM QSO_4. 10 ml samples were sipped from 50 ml beakers and expectorated. Subjects were required to describe, not identify, the stimuli in their own words; no taste words were supplied to subjects so as not to bias them by suggestion (O'Mahony and Thompson, 1977). Subjects were also required to match any of the samples which had the same taste. All responses were given verbally.

Results

3.1 Experiment 1

32/45 (71.1%) of FST students and 56/100 (56%) non-FST students correctly categorised all 12 stimuli. This 100% criterion was maintained as the criterion for a 'correct' subject throughout the study. Significantly more FST students were correct than incorrect (chi square, $p < 0.01$) which was not so for non-FST students. Thus FST students, probably due to their specialised courses in the chemical senses and experience on taste panels, were a substantial and practiced group as far as initial taste naming was concerned and were treated separately. Of the non-FST students, there was not a substantial number with an academic background that might seem to bias results.

The percentages of quality responses reported for the various stimuli are given in Table II. Because some stimuli were presented more than once, an alternative analysis would be to note the percentage of subjects giving each response; this, however, does not give a substantially different picture from that shown in Table II. Although many misnamings occurred, the most common were sour-bitter confusions with a greater tendency to call citric acid 'bitter'. Of the total number of non-FST subjects 21% called 20 mM citric acid 'bitter', 16% called 25 mM citric acid 'bitter', 17% called 200 μM quinine sulphate 'sour', at least once. For FST students these values were 11.1%, 8.9% and 11.1% respectively. There was also a lesser tendency to confuse 'sour', and 'salty' as well as a tendency to call citric acid 'sweet', although not vice versa.

Of those who agreed to return for retesting 3-4 weeks later there were 28 correct non-FST (13M, 15F, 17-31 yrs), 28 correct FST (8M, 20F, 20-27 yrs) and 22 incorrect both non-FST and FST (10M, 18F, 18-26 yrs) subjects. Of those formerly correct 26/28 (92.9%) FST subjects remained correct, significantly more than fell into error (chi square, $p < 0.001$), while only 19/28 (67.9%) non-FST subjects remained correct, this number being not significantly greater than those who fell into error (chi square, $p > 0.05$). Thus FST students when correct tended to remain correct. Of those formerly incorrect only 6/22 (27.3%) remained incorrect, significantly more became correct (chi square, $p < 0.05$). Of the 6 consistently incorrect

Table II. Percentage of quality responses to the stimuli presented initially in Experiment 1.

Subjects	Solutions (mM)	Qualities					
		salty	sour	bitter	sweet	peppermint	tasteless
non-FST (N = 100)	300 NaCl	97.5	1.5	0.5	0.5	—	—
	350 NaCl	97.5	0.5	1.0	—	—	1.0
	400 NaCl	99.0	—	—	1.0	—	—
	20 Citric Acid	2.0	81.0	14.0	1.5	1.5	—
	25 Citric Acid	1.0	82.0	16.0	1.0	—	—
	0.200 Quinine Sulphate	—	9.0	90.0	—	0.3	0.7
	300 Sucrose	—	—	—	100.0	—	—
FST (N = 45)	300 NaCl	100.0	—	—	—	—	—
	350 NaCl	98.9	1.1	—	—	—	—
	400 NaCl	95.6	2.2	—	2.2	—	—
	20 Citric Acid	1.1	92.2	5.6	1.1	—	—
	25 Citric Acid	—	91.1	8.9	—	—	—
	0.200 Quinine Sulphate	—	4.4	91.9	—	2.2	1.5
	300 Sucrose	—	—	—	100.0	—	—

FST denotes a student whose specialisation is Food Science and Technology.

subjects, 3 needed one learning trial and 3 needed two learning trials before achieving the criterion of 100% correct application of adjectives.

The data were also analysed using two less strict criteria for correctness than no errors, namely 0-1 errors and 0-2 errors. Such analyses reduced the numbers of incorrect subjects but did not alter the trends already noted except that differences between FST and non-FST subjects vanished.

3.2 Experiment 2

40/100 (40%) of the subjects correctly categorised all 12 stimuli; significantly more were incorrect (chi square, $p < 0.05$). It should be noted that there was a smaller proportion of errorless subjects than in Experiment 1. The percentage of quality responses reported for the stimuli are given in Table III; a higher level of misnaming occurred than for Experiment 1 with the tendency to call citric acid 'bitter' most common followed by calling QSO_4 'sour'. There was also a lesser tendency to confuse 'sour' and 'salty', as well as a tendency to call citric acid 'sweet' but not vice versa. Also of note is the increased frequency of 'tasteless' responses indicating that although stimuli could be distinguished a 'tasteless' description was sometimes appropriate.

Of the 21 initially correct subjects (9M, 12F, 18-28 yrs), 17 (81%) remained correct on retasting 3-4 weeks later, significantly more than erred (chi square, $p < 0.01$). Of those initially incorrect (8M, 19F, 17-35 yrs) 17/27 (63%) remained incorrect, not significantly more than became correct. The 17 consistently incorrect subjects, more than for Experiment 1, were given learning trials; 13 (76.5%) attained 100% criterion performance after one trial, 3 (17.6%) after two trials, one (5.9%) after three trials.

Reanalysis of the data using the more lax criteria used in Experiment 1 increased the number of 'correct' subjects but did not substantially alter the trends already noted.

Table III. Percentage of quality responses to the stimuli presented initially in Experiment 2.

Subjects	Solutions (mM)	Qualities					
		salty	sour	bitter	sweet	peppermint	tasteless
non-FST	100 NaCl	88.0	2.0	1.0	1.0	1.0	7.0
(N = 100)	150 NaCl	92.0	1.0	4.0	1.0	—	2.0
	200 NaCl	94.0	4.0	—	1.0	1.0	—
	250 NaCl	100.0	—	—	—	—	—
	300 NaCl	99.0	—	1.0	—	—	—
	10 Citric Acid	3.0	74.0	19.0	3.0	1.0	—
	15 Citric Acid	2.0	74.0	20.0	3.0	1.0	—
	20 Citric Acid	2.0	80.0	13.0	4.0	1.0	—
	0.040 Quinine Sulphate	—	7.0	81.0	—	1.0	11.0
	0.045 Quinine Sulphate	—	9.0	80.0	—	4.0	7.0
	0.050 Quinine Sulphate	—	5.0	86.0	—	1.0	8.0
	100 Sucrose	1.0	—	—	97.0	1.0	1.0

non-FST denotes subjects whose specialisation was not Food Science and Technology

3.3 Experiment 3

In this experiment, the procedure indicated that the four groups of stimuli, 'sweet', 'sour', 'salty', 'bitter', could not be readily distinguished from each other: only 2 (4%) subjects made no incorrect groupings. Grouping two separate modalities in one group was regarded as an error, splitting a modality into two or more groups (strong bitter, weak bitter) was not. The most common error was to group the three lowest concentrations of QSO_4 with either 50 mM NaCl or 15 mM sucrose. This experiment was not continued further although a second group of 50 subjects tasted and named stimuli at these indistinguishable concentrations and had the expected high error rate; only one subject applied adjectives correctly.

3.4 Experiment 4

19/27 (77.8%) subjects correctly categorised all 12 stimuli until confronted by the lowest concentration stimuli in the final trial, while 2 (7.4%) subjects made no errors at all. Misnamings occurring in the final part are to be expected from the lack of distinguishability noted in Experiment 3; all misnamings involved calling stimuli 'tasteless'. A further 6 subjects made errors not only at the low concentrations but also at the medium concentrations used in Experiment 2. These were mainly sour-bitter confusions (QSO_4 called 'sour' 40%; citric acid called 'bitter' 33.3%). Other errors involved calling sucrose 'tasteless' (6.7%), QSO_4 'tasteless' (13.3%) and NaCl 'bitter' (6.7%). Of these six subjects, two even made errors at the highest concentration (all sour-bitter confusions). Thus generally, subjects did not make errors until they tasted the lowest (indistinguishable) concentrations. Of those subjects who did make errors at other concentrations, the majority tendency was to make errors at only the medium concentrations.

3.5 Experiment 5

39/100 subjects correctly categorised all 12 stimuli; significantly more were incorrect than correct (chi square, $p < 0.05$). This was a comparable proportion to that in Experiment 2. The percentage of quality responses reported for the stimuli are given in Table IV; the total proportion of responses that were inappropriate in

Table IV. Percentage of quality responses to the stimuli presented initially in Experiment 5.

Subjects	Solutions (mM)	salty	sour	bitter	sweet	peppermint	tasteless
npn-FST	100 NaCl	93	2	2	1	—	2
(N = 100)	150 NaCl	97	—	1	1	—	1
	200 NaCl	94	4	1	1	—	—
	250 NaCl	96	2	1	—	1	—
	300 NaCl	96	3	1	—	—	—
	10 Citric Acid	1	74	17	7	1	—
	.15 Citric Acid	4	72	15	6	3	—
	20 Citric Acid	1	73	22	4	—	—
	0.040 Quinine Sulphate	—	15	66	1	—	18
	0.045 Quinine Sulphate	—	16	73	—	1	10
	0.050 Quinine Sulphate	—	10	77	1	1	11
	200 Sucrose	—	—	1	93	4	2

non-FST denotes subjects whose specialisation was not Food Science and Technology

Experiments 2 and 5 were similar (14.8%, 19.5% respectively). The most common error was calling citric acid 'bitter' followed by calling QSO_4 'sour'; there was also a tendency to confuse 'sour' and 'salty', as well as a tendency to call citric acid 'sweet' but not vice versa. There were also a substantial number of 'tasteless' responses for QSO_4.

Of the 10 subjects recalled for a second tasting for the learning trial four made no mistakes on re-testing. After the learning trial the 6 remaining subjects made no sour-bitter confusions. All errors of categorising sucrose were corrected although one subject categorised NaCl as 'bitter'.

Regarding the testing of stimuli by 10 recalled subjects at two concentrations, subjects made significantly fewer total errors at higher concentrations (Wilcoxon, $p < 0.02$), despite there being no learning trials. In both cases the most common error was naming citric acid 'bitter' then naming QSO_4 'sour'. Salty-sour confusions were made but citric acid was never named 'sweet'. On passing from lower to higher concentration solutions the number of tasteless descriptions decreased.

3.6 Experiment 6

Of the 297 subjects 76 (25.6%) made errors on their first trial; the percentage of quality responses reported for the stimuli are given in Table V. Again the most common misnaming was to call citric acid 'bitter' followed by calling QSO_4 'sour'. There was also the sour-salty confusion noted in previous experiments as well as a tendency to call citric acid 'sweet'. Of the 76 subjects who initially made errors, 23 made errors on immediate retesting (not significantly less than spontaneously recovered, $p > 0.05$) and were given learning trials; 15 subjects needed one trial, 7 needed two and one needed three. 20 of these subjects were retested 3-4 weeks later and 17 were found to make no errors. Of the three who did make errors, only one made an error the same as 3-4 weeks earlier.

3.7 Experiment 7

10 subjects reported that all stimuli tasted different while 9 stated that two stimuli tasted the same and 3 stated that three stimuli tasted the same.

3.8 Experiment 8

17 subjects reported that all stimuli tasted different while 9 stated that two stimuli tasted the same and one stated that three stimuli tasted the same.

3.9 Experiment 9

Although some subjects named the stimuli incorrectly, all reported that the two

Table V. Percentage of quality responses to the stimuli presented initially in Experiment 6.

Subjects	Solutions (mM)	salty	sour	bitter	sweet
non-FST	400 NaCl	97.3	0.7	1.7	0.3
(N = 297)	300 Citric Acid	2.2	85.7	9.8	2.3
	0.2 Quinine Sulphate	1.7	5.1	93.3	—
	300 Sucrose	0.3	0.3	—	99.3

non-FST denotes subjects whose specialisation was not Food Science and Technology

stimuli were distinguishable.

3.10 Experiment 10

Considering the pair comparisons in ascending concentration, the numbers of subjects who responded correctly were 11 (36.7%), 10 (33.3%), 14 (46.7%) and 14 (46.7%). Others either identified incorrectly or spontaneously reported that they could not tell any difference.

3.11 Experiment 11

No subjects reported that 'sour' and 'bitter' stimuli tasted the same as each other but 22 (76%) subjects reported the HCl and citric acid to taste the same and 21 (72%) subjects the two QSO_4 stimuli. The descriptions given to these stimuli reflected the confusions although QSO_4 was reported as 'tasteless' or 'water' for 66% of the descriptions for 13 μM QSO_4 and 59% for 8 μM QSO_4.

Discussion

There are six main points to note from these experiments. Firstly, errors of application of taste adjectives, including sour-bitter confusions, can be corrected merely by defining taste adjectives by providing the appropriate taste sensations (Experiments 1, 2, 5 and 6). Secondly, not all subjects remained consistent; some who misapplied adjectives initially did not necessarily do so when retested before any learning trials. Conversely, some of those initially correct erred later (Experiments 1, 2, 5 and 6). Thirdly, while stimuli were still distinguishable, fewer naming errors occurred for stimuli at higher concentrations than lower concentrations (Experiments 1, 2, 4, 5 and 6). Fourthly, if stimuli were too dilute, they could not always be distinguished from each other (Experiment 3); partial replication of the work of prior investigators (Experiments 7-11) indicates that, for some studies, the stimuli were not always distinguishable. The fifth point concerns sour-bitter confusions; calling citric acid 'bitter' was more common than calling QSO_4 'sour'. Lastly, lesser though consistent tendencies to confuse 'sour' and 'salty' and to call citric acid 'sweet' were also noted. These points will be discussed in more detail.

The main result that errors in the application of taste adjectives can be corrected merely by giving subjects the appropriate standard, occurred for several variations in method and would seem to be fairly stable, at least over a few weeks, and work fairly substantially when only some of the adjectives were defined (of Experiments 1, 2, 5 and 6). Thus it would seem that correct application of taste adjectives is more a matter of learning (Robinson, 1970) and that confusions such as the sour-bitter confusion are not physiological analogues of color blindness (Meiselman and Dzendolet, 1967).

It may then be asked why the correction procedure used by Meiselman and Dzendolet did not have the success that it had in the present study. There are several differences between the two studies. Their procedure did not involve definition of the standard, it merely informed subjects how the majority of subjects would respond while insisting that subjects use their own opinions. It may be hypothesised that this less forceful procedure would have less forceful results. The high number of 'no taste' responses in their study and the fact that in the present study a high proportion of subjects reported stimuli of comparable concentration tasted the same (Experiments 7 and 8) would suggest that not all subjects could distinguish between the stimuli. For such subjects, no correction procedure would have worked. Incidentally,in relation to this, it is worth noting that all the subjects in Experiment 9 reported that the higher concentration stimuli used by Robinson were distinguishable.

Thus it is possible that some of Meiselman and Dzendolet's subjects could not distinguish all the stimuli but it is likely that others could. Of these latter subjects, it is possible that the mild correction procedure was not always sufficient to correct all mistakes. Their study does indicate a small degree of improvement for HCl, NaCl and sucrose stimuli but, mysteriously, things got worse for QSO_4.

Only Robinson (1970) and Meiselman and Dzendolet (1967) were wholly concerned with application of taste adjectives. Gregson and Baker's (1973) study was more concerned with the fact that subjects could give both 'sour' and 'bitter' intensity ratings for citric acid or QSO_4. They reasoned that this may be caused by a reversion to aqualitative scaling. However a further 'suggestion' or 'experimenter pleasing' effect may also be hypothesised. Subjects when instructed to rate bitter substances for sourness (and vice versa), especially if stimuli are sufficiently dilute that their taste quality is not clear, may well give appropriate responses; whether these responses indicate actual perceptions remains a matter for debate. The low distinguishability of some of their 'sour' and 'bitter' solutions, seen in Experiment 10, would lend support to this argument.

McAuliffe and Meiselman's (1974) study was not really concerned directly with correcting the misapplication of adjectives. It was concerned with matching of stimuli by subjects who had previously shown inability to correctly name 'sour' and 'bitter' stimuli. They noted that familiarity rather than feedback was more important for improvement in stimulus matching. This is not surprising because stimulus matching does not depend to such a great extent on the correct use of taste adjectives. Experiment 11 would suggest that the 'sour' and 'bitter' stimuli were sufficiently different to be distinguished, although within modality confusion occurred. The high number of 'tasteless' or 'water' responses for QSO_4, in both studies, suggests that QSO_4 may have been distinguished from citric acid by its lack of taste rather than its bitter taste. Thus it may be hypothesised that in the prior screening, the subjects had not necessarily applied the adjectives 'sour' and 'bitter' incorrectly, they may merely have reported that the QSO_4 was 'tasteless'. Correct performance in a taste matching task for these subjects would have involved not only the matching of perceptible sensations (or lack of them) but also learning to identify the presence of near threshold sensations for low concentration stimuli.

Experiments 3, 7, 8, 9 and 11 required subjects to make 'same-different' judgments which are subject to criterion variation (Green and Swets, 1966), unlike the forced choice procedure used in Experiment 10. Thus, the judgments of distinguishability probably underestimate the performance that may have been encountered in the studies of Meiselman and Dzendolet (1967), Robinson (1970) and McAuliffe and Meiselman (1974) because stimuli given a 'same' response may sometimes be distinguished. However, in studies where indistinguishability of stimuli occurred, it was frequent enough to sound a cautionary note. Further, these experiments certainly vindicate the choice of higher concentrations than those used in prior studies, to test adjective application skills in Experiments 1, 2, 5 and 6.

A strict criterion of correctness (100%) was adopted in these studies, a criterion far stricter than that adopted in some prior studies. The adoption of a more lax criterion does not substantially alter the results except that students specialising in Food Science and Technology (FST) no longer emerge as a group better skilled in applying taste adjectives. This would indicate that a high proportion of FST students make no errors but many non-FST students only make one or two errors. In the light of this, it could still be argued that the study of Food Science and Technology at U.C.D. may expose students, through taste panels or courses of study, to taste standards. Psychology students at Bristol do not have such a background and so were not considered to be a special group. This indicates the importance of identifying specialised background experience when testing naive subjects. There was too little evidence to identify any further special groups of subjects.

An effect not noted before was the spontaneous recovery of many subjects who had initially made errors and the erring of many subjects who had previously been correct (Experiments 1, 2, 5 and 6). This would suggest an instability over time for taste confusions; the reasons for this are not obvious. For some of those subjects who showed spontaneous recovery, it could be hypothesised that they were not fully attending initially, perhaps due to lack of motivation or unfamiliarity with taste testing. Only on retesting were they sufficiently motivated or at ease to perform to their full potential. This would be supported by the fact that initially correct FST students had a far greater tendency to remain correct, because FST students tend to be more sophisticated in taste testing. Those who were initially correct and later erred may have only been correct by chance. However, at this point there is insufficient evidence to draw any firm conclusions.

It is interesting to note that despite the fact that stimuli could be clearly distinguished, fewer misnamings occurred at higher concentrations (Experiments 1, 6) than lower ones (Experiments 2, 5). Testing this further using a related samples design (Experiment 4), a few subjects were noted who made more errors when changing from higher to lower concentration stimuli while the reverse happened on passing from lower to higher concentrations (Experiment 5). This concentration effect is not the same as an inability to distinguish low concentration stimuli. It may be that the subjects' concepts of 'sweet', 'sour', 'salty' and 'bitter' correspond more with the experimenter's at higher concentrations of the stimuli than at lower concentrations; there is no reason to assume that the concepts should coincide at all concentrations. For example, the sensations included in the subject's concept of 'bitter' may include those experienced when tasting low concentrations of citric acid; this would be seen by the Experimenter as a sour-bitter confusion.

Misapplication of the adjective 'sour' to QSO_4 and 'bitter' to citric acid in Experiments 1, 2, 5 and 6 demonstrates the sour-bitter confusion. Of the 1629 responses obtained for sour and bitter stimuli in these experiments, 13.3% involved calling citric acid (10-300 mM) 'bitter', while 7.7% involved calling QSO_4 (40-200 μM) 'sour'. Such confusions with the predominance of sour stimuli being named 'bitter', occurred in prior studies (Meiselman and Dzendolet, 1967; Robinson, 1970;

McAuliffe and Meiselman, 1974) for lower concentrations of QSO_4 and citric acid and also when HCl was used as a sour stimulus and KCl as a bitter stimulus (Meiselman and Dzendolet, 1967; McAuliffe and Meiselman, 1974). However, comparison with studies using different stimuli must be made with caution because their tastes may vary in quality from the tastes of stimuli used here.

Although all types of confusions tended to be made from time to time, there was a consistent tendency in all experiments for NaCl to be called 'sour' and citric acid to be called 'salty'. This sour-salty confusion had been noted in prior studies; Meiselman and Dzendolet (1967) noted a tendency to call NaCl 'sour' and HCl 'salty' while Robinson (1970) noted a tendency to call citric acid 'salty'. Myers (1904) also reported incorrect application of the equivalent words for 'sour' and 'salty' to sour and salty stimuli in other cultures, for example: Torres Straits islanders. A further consistent tendency noted in the experiments was to call citric acid 'sweet' (but not sucrose 'sour'); this was also noted by Robinson (1970). Meiselman and Dzendolet (1967) also noted 'sweet' and 'sour' misnaming for HCl and sucrose respectively. The sour-salty and the sweet citric and 'confusions' noted in Experiments 1, 2, 5 and 6, were, although consistent, relatively minor compared to the sour-bitter confusions.

In the experiments reported here, more errors occurred in the use of 'sour' and 'bitter' than 'sweet' and 'salty'. It could be hypothesised that because subjects probably have more experience in their culture of salty and sweet foods than sour or bitter foods, their definitions of 'salty' and 'sweet' would have the chance to develop into something like those required in this experiment; subjects who have eaten little sour or bitter food would hardly have the experience to develop concepts of 'sour' or 'bitter'. As a second hypothesis, it may be argued that because most subjects will have had experience with table salt and sugar in their pure form, their application of the adjectives 'salty' and 'sweet' would have been refined to correspond fairly well with the definitions in the experimental context based on NaCl and sucrose solutions. Most subjects will have tasted table salt and sugar in their pure form and the present study shows that one exposure can be enough to learn to use the adjective correctly. Because citric acid and QSO_4 are rarely, if ever, eaten in pure form, the everyday definitions of 'sour' and 'bitter' could be erroneous. Confusion could be generated, for example, by the common habit in English of describing lemons as bitter; this effect is reinforced by the names of soft drinks such as 'Bitter Lemon'. An additional confusion may also occur with the application of the term 'bitter' to unpleasant stimuli. Whether either one, both or none of these hypotheses are correct is a matter for further experimentation.

There would seem to be no obvious reason for a sour-salty confusion based on food naming; perhaps some similarity in sensation may be noted. It may be hypothesised that the naming of citric acid as 'sweet' is due to the subjects' concept of 'sour' being related to citric fruits which are both 'sour' and 'sweet'. Citric acid may be tasted as 'fruit-like' and then called 'sweet'.

It cannot be assumed that the subjects' taste concepts, abstracted from their experience of foods, are necessarily as consistent as those defined by the experimenter. The high number of errors with lower concentration stimuli could be due to the higher correspondence between the layman's and experimenter's use of the adjectives for strong tastes than for weak tastes. Strong black coffee may warrant being called 'bitter' by the layman but not weak coffee, even though a bitter taste is perceived. This could result in application of 'bitter' to strong QSO_4 but not QSO_4 at lower concentrations. Context may also be important, the same 'sour' sensation may be labelled 'sour' in the context of a lemon but 'bitter' in the context of a soft drink.

Following such hypotheses, it can be argued that the names applied to the stimuli found in an experimental context by naive subjects would depend on the language structure of the community. This, in turn, would depend on the foods eaten by that community and 'errors' would be perpetuated from generation to generation. Such 'errors' are merely a difference in application of terms by scientist and layman, an arbitrary effect noted in many branches of science. Regardless of the exact mechanisms of the development of taste concepts for untrained judges, it is important to note that the assumption that the experimenter and subject will apply the same adjectives to taste solutions, without the prior use of standards, is dangerous. The sour-bitter confusion as so defined, would seem to be, in the absence of pathological taste deficits, merely a lack of routine pre-experimental training.

Acknowledgements

The authors would like to thank Marla Nelson, Nina Ballart, Betty Medeiros, Roberta Lew, Susan Carleton and Cecilia Lum for their assistance.

* Present address: Department of Psychology, University of Bristol, Bristol BS8 1HH, England.

References

Amerine, M.A., Pangborn, R.M. and Roessler, E.B.: 1965, 'Principles of Sensory Evaluation of Food', Academic Press, London.

Caul, J.F.: 1957, 'The profile of flavor analysis', *Adv. Food Res.* 7, 1-40.

Caul, J.F., Cairncross, S.E. and Sjöström, L.B.: 1958, 'The flavor profile in review', in Flavor Research and Food Acceptance, Chapman & Hall, London. pp 65-74.

Chamberlain, A.F.: 1903, 'Primitive taste-words', *Am. J. Psychol.* 14, 146-153.

Green, D.M. and Swets, J.A.: 1966, Signal Detection Theory and Psychophysics, Wiley, New York.

Gregson, R.A.M. and Baker, A.F.H.: 1973, 'Sourness and bitterness: confusions over sequences of taste judgements', *Brit. J. Psychol.* 64, 71-76.

McAuliffe, W.K. and Meiselman, H.L.: 1974, 'The roles of practice and correction in the categorization of sour and bitter taste qualities', *Perception and Psychophysics* 16, 242-244.

Meiselman, H.L. and Dzendolet, E.: 1967, 'Variability in gustatory quality identification', *Perception and Psychophysics* 2, 496-498.

Myers, C.S.: 1904, 'The taste-names of primitive peoples', *Brit. J. Psychol.* 1, 117-126.

O'Mahony, M.: 1972a, 'The interstimulus interval for taste: 1. The efficiency of expectoration and mouthrinsing in clearing the mouth of salt residuals'. *Perception* 1, 209-216.

O'Mahony, M.: 1972b, 'The interstimulus interval for taste: 2. Salt taste sensitivity drifts and the effects on intensity scaling and threshold measurement', *Perception* 1, 217-222.

O'Mahony, M.: 1973, 'A note on the reliability of subjective judgement in discerning when the mouth is clear of salt taste stimuli', *Perception & Psychophysics* 14, 437-439.

O'Mahony, M. and Dunn, M.: 1974, 'Do sensitivity drifts occur for stimuli other than sodium chloride?: A preliminary investigation', *Perception* 3, 213-220.

O'Mahony, M and Godman, L.: 1974, 'The effect of interstimulus procedures on salt taste thresholds',

Perception & Psychophysics **16**, 459-465.
O'Mahony, M and Wingate, P.: 1974, The effect of interstimulus procedures on salt taste intensity functions', *Perception & Psychophysics* **16**, 494-502.
O'Mahony, M. and Thompson, B.: 1977, 'Taste quality descriptions: Can the subject's response be affected by mentioning taste words in the instructions?', *Chem. Senses & Flavor* **2**, 283-298.
O'Mahony, M., Autio, J., Heintz, C. and Goldenberg, M.: 1978, 'Taste naming by preschool children compared to colour naming: Preliminary examination', *IRCS Med. Sci.* **6**, 208.
Pangborn, R.M.: 1961, in discussion for Pilgrim, F.J., 'Interactions of suprathreshold taste stimuli', in Physiological and Behavioral Aspects of Taste, University of Chicago Press, Chicago, p. 72.
Robinson, J.O.: 1970, 'The misuse of taste names by untrained observers', *Brit. J. Psychol.* **61**, 375-378.
Stone, H., Sidel, J., Oliver, S., Woolsey, A. and Singleton, R.C.: 1974, 'Sensory evaluation by quantitative descriptive analysis', *Food Technol.* **28**, 24-34.
Williams, A.A.: 1975, 'The development of a vocabulary and profile assessment method for evaluating the flavour contribution of cider and perry aroma constituents', *J. Sci. Food Agric.* **26**, 567-582.