

Perception of caffeine and its effects: Laboratory and everyday abilities

JOSEPH P. BLOUNT

University of Minnesota, Morris, Minnesota

and

W. MILES COX

*Richard L. Roudebush Veterans Administration Medical Center
and Department of Psychiatry, Indiana University School of Medicine
Indianapolis, Indiana*

Three studies investigated the ability to perceive caffeine via taste or bodily effects, under different concentrations and in different brands of coffee. In Experiment 1, 40 subjects ingested capsules containing 0, 200, 400, or 600 mg of caffeine. After a 90-min delay, they made magnitude estimates and category ratings of the caffeine ingested. Results showed significant linear trends between actual and judged caffeine levels. In Experiment 2, 89 subjects tasted samples of coffee with concentrations corresponding to the doses of Experiment 1. The estimated levels of caffeine again showed a significant linear relationship to actual levels. In Experiment 3, 63 subjects tasted 24 samples of coffee and judged which of two brands each sample was and whether each was caffeinated or decaffeinated. Brand, but not caffeination, was identified with better than chance accuracy. It was concluded that the ability to perceive caffeine in coffee is lost somewhere between 200 mg/cup and 50 mg/cup.

Knowledge about the sensory and perceptual properties of caffeine has theoretical importance as well as practical implications for both consumers and producers of food and beverages. Consumers must balance perceived caffeine intake, desired effects (such as arousal), and possible harmful effects. Much of the caffeine research of the 1970s reported deleterious physical effects (e.g., Jick et al., 1973, for heart disease; Ritchie, 1975, for peptic ulcers; Manber, 1976, for a popular summary; "Caffeine: How to Consume Less," 1981, and Brody, 1982, for advice on reducing caffeine intake). Consequently, over the past few years, there has been a surge of public interest in the possible harmful effects of caffeine and concern about whether particular foods and beverages contain caffeine. Interest has also increased in whether caffeine-free products sacrifice something in taste. Such interest makes caffeine an apt substance for studying such general perceptual problems as (1) the perception of a substance through its effects, (2) the perception of a flavor in complex mixtures, and (3) the relative salience of flavors in a mixture.

This research was supported in part by a grant-in-aid from the Graduate School of the University of Minnesota to Joseph P. Blount and W. Miles Cox. We gratefully acknowledge the assistance of Craig Bowden, Chris Dallager, Mike Johnson, and Lynn Sax in the data collection. Joseph Blount is now at the University of New Hampshire, Durham, NH 03824.

Requests for reprints should be sent to Miles Cox at Psychological Service (116B), Veterans Administration Medical Center, 1481 West 10th St., Indianapolis, IN 46202.

However, throughout the 1970s most of the research continued to be in the medical field. This research failed to substantiate the early findings of harmful effects of caffeine (e.g., Cohen & Booth, 1975, regarding peptic ulcers; Dawber, Kannel, & Gordon, 1974, regarding coronary heart disease). Only recently has research been expanded to include the psychological effects of caffeine (Blount & Cox, 1982; Cox & Blount, 1984). The psychological research has been concerned with the sensory-perceptual, psychopharmacological, and placebo effects of caffeine, interactions between caffeine and personality, induced changes of mood, and effects of caffeine on performance in cognitive and motor tasks.

Coffee is the most widely used product containing caffeine. Coffee appears to be consumed both for the bodily effects it provides and for its taste. The major caffeine-related reasons people list for drinking coffee, are "the lift of energy" it provides, "to stay awake," "to get started in the morning," and "its taste" (Blount & Cox, 1983). These subjective reports imply underlying perceptual abilities based upon flavor and bodily arousal. On the other hand, experimenters' use of decaffeinated products as placebos presupposes the lack of such perceptual abilities (that could nullify the use of the placebo). To resolve this issue, we designed a series of three experiments to study the extent of people's ability to detect the presence and concentration of caffeine and the pharmacological effects of various dosages of caffeine.

EXPERIMENT 1

The stimulation from caffeine that people perceive may result from both psychological expectancy and pharmacological effects. In Experiment 1, psychological expectancies were held constant, and there were no gustatory, olfactory, visual, or tactual cues to the dosage of caffeine, permitting us to assess the accuracy with which subjects can detect the pharmacological effects of caffeine. It was hypothesized that secretion of gastric acids, relaxation of smooth muscles, changes in heart rate, excitation of the central nervous system, and other bodily clues might contribute to the perception of caffeine even though people might not be consciously aware of each individual bodily change. (See Teghtsoonian, Becker, & Edelman, 1981, for another example of scaling bodily sensations.)

In Experiment 1 we did not use extensive training and/or repeated trials because the goal was to characterize performance that might be expected in everyday life. Thus, each subject judged a single sample and estimated the magnitude of the caffeine effect. This procedure contrasts with the usual indirect methods of obtaining functional relationships between physical stimuli and sensory responses on the basis of just noticeable differences (see Schutz & Pilgrim, 1957, for caffeine taste; Jesteadt, Wier, & Green, 1977, for loudness; McBride, 1983, for sucrose taste).

Method

Subjects. The subjects were 15 males and 25 females who were recruited from university classes and received \$4 for participating. (A few subjects served gratis at the experimenter's request.) Potential subjects were screened to eliminate those with heart disease or allergic reactions to caffeine and females who might be pregnant. Subjects were not required to be coffee drinkers. The subjects were divided randomly into four groups with approximately equivalent male/female ratios.

Materials. In a double-blind procedure, two size-0 gelatin capsules were prepared for each subject and placed in an envelope identified only with a subject number. The capsules were filled with either powdered sugar (the no-caffeine condition) or a mixture of powdered sugar and 200, 400, or 600 mg reagent caffeine. A chemist's analytical balance specified caffeine weight to ± 5 mg. By one of the common standards, 100 mg per 5-oz cup, these values represent 2, 4, or 6 cups. To a heavy coffee drinker who brews his or her own drip coffee and uses a large mug, these values represent 1, 2, or 3 "cups" (i.e., taking strength of coffee and size of cup into account). Because negligible amounts of caffeine are excreted in the urine and its biologic half-life averages 3.5 h (Axelrod & Reichenenthal, 1953), caffeine accumulates in the bodies of heavy drinkers. The 600-mg dose was chosen to represent blood concentrations heavy drinkers may reach (although some users may achieve that level gradually rather than through a single oral dose).

Subjects completed a caffeine habits questionnaire on which they provided detailed information about their habitual use of caffeine. They reported the quantity and frequency with which they used various products containing caffeine (e.g., caffeinated and decaffeinated coffee, tea, all brands of soft drinks, chocolate products, over-the-counter and prescription medications) and indicated the settings in which these products were consumed. In addition, subjects reported their usual time of arising, and the time that they arose for the experiment.

Response scales. Subjects were asked to make magnitude estimates in terms of equivalent "number of cups of coffee." The sub-

ject's memory for his or her usual cup of coffee thus served as the modulus or standard, and followed Poulton's (1979) suggestion that responses are less biased if they are linked to the stimuli by well-known rules or familiar physical units. Our subjects were asked to specify the type of coffee they had in mind (instant, brewed, or drip) so that the familiar units could be converted to standard physical units.

We also used a second judgmental task, a category scale comprising five verbal descriptors: (0) no caffeine at all, (1) below-average amount of caffeine, (2) average amount, (3) above-average amount, and (4) large amount. The use of a verbal descriptor for each category is in accord with Bendig and Hughes' (1953) finding that verbal anchors increase the amount of information transmitted by a scale.

Poulton (1979) argued that a researcher can avoid nearly all the possible biases in perceptual scaling by asking unpracticed observers to judge only a single stimulus, with unbiased instructions and no demonstrations. Each of our subjects made only one magnitude judgment and one category rating.

Length of delay. A 90-min delay was used between ingestion of caffeine and judgment of quantity ingested. (This delay interval was chosen after conducting a pilot experiment with a 45-min delay. Pilot subjects made inaccurate judgments, and many of them spontaneously returned to report that they felt the maximal effects only after leaving the experiment. Because the decay of caffeine effects is slower than the rise, it is more conservative to overestimate than to underestimate the period for biological diffusion. Therefore, a longer delay of 90 min was chosen.)

Procedure. Subjects were instructed not to consume any caffeine for 8 h prior to the experiment, and were provided with a list of caffeinated products to avoid; no other restrictions were placed on food or beverage intake. When they arrived for the experiment at 8:00 a.m. on a weekday, subjects were told that their experimenter didn't know which doses would be received by which individuals. They signed consent forms and affidavits stating that they had, in fact, abstained from caffeine for the previous 8 h. They were given their assigned envelopes and ingested the capsules with ad lib water, then completed the caffeine habits questionnaire. Each subject was taken to a small, quiet, windowless cubicle and allowed to study alone with reading material she or he had been instructed to bring. After 90 min, the experimenter returned and asked the subject (1) to estimate on a five-category scale the amount of caffeine that she or he had ingested, (2) to estimate the number of cups of coffee that she or he thought would equal the amount of caffeine in the capsules, and (3) to describe any effects she or he felt from the capsules. The subject was then dismissed. (If subjects wished to know the actual dosage of caffeine they had ingested, this information was later provided by a second experimenter, who had assigned subject numbers to dosage levels.)

Results

Because of the small number of data points (10 per group), three analyses were performed to ensure that randomization had given approximately equivalent groups. There were no significant differences among groups in what time (Mean = 7:54 a.m., SD = 73 min) they arose ($F < 1$), in how much earlier (Mean = 45 min, SD = 60 min) this was than usual ($F < 1$), and in how much caffeine (Mean = 207 mg, SD = 224 mg) they ingested per day [$F(3,36) = 1.2, p > .33$]. The four groups were equivalent in factors that might influence their judgments—arousal, disruption, and acquired sensitivity or tolerance (Colton, Gosselin, & Smith, 1968; Goldstein & Kaizer, 1969; Revelle, Humphreys, & Gilliland, 1980). A further preliminary analysis showed that caffeine consumption could not be used as a covariate to increase experimental

precision because it did not have significant linear correlations with the dependent variables ($p > .99$ for magnitude estimates; $p > .42$ for category ratings). Accordingly, only analyses of variance were used.

Magnitude estimates were analyzed in two ways. In the first analysis, the simple number of cups was used. In the second, subjects' estimated number of cups was multiplied by the actual caffeine content of the designated type of coffee in order to arrive at an "estimated number of milligrams." We had anticipated that the second analysis would improve accuracy. In fact, both measures produced essentially identical patterns of means, and analyses of variance showed significant effects in both. We noted, however, that the "estimated milligrams" data showed greater variance than the "cups" data. Upon further reflection, it was decided that the "milligrams" data give a false sense of accuracy: chemists' assessments vary widely, the size of cup is not taken into account, basing calculations on one reported kind of coffee does not ac-

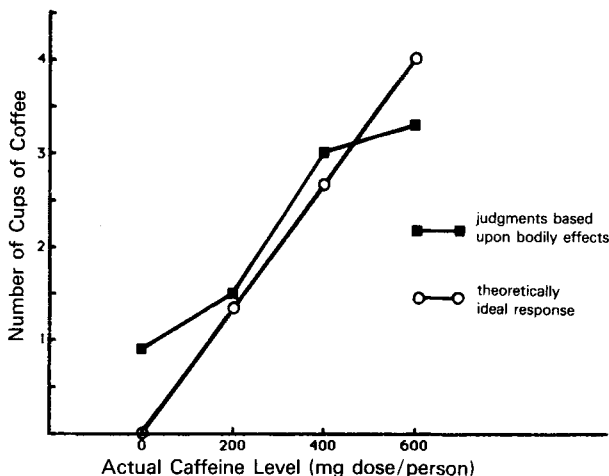


Figure 1. Mean magnitude estimates of caffeine level with judgments made 90 min after ingestion, Experiment 1.

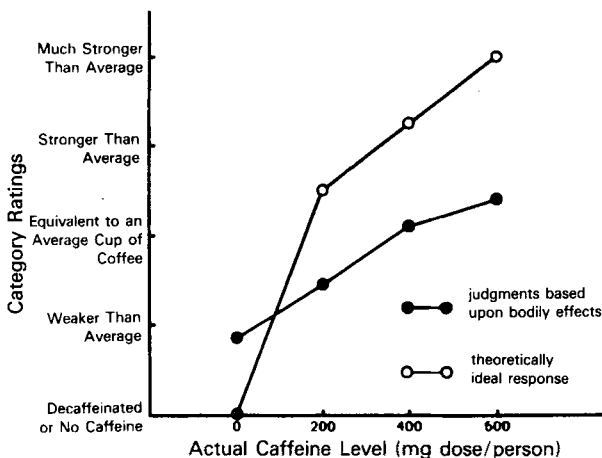


Figure 2. Mean category ratings of caffeine level with judgments made 90 min after ingestion, Experiment 1.

curately represent users whose experience involves several kinds, and individual methods of preparation cause great variation within one commercial form of coffee (Gilbert, 1976). Therefore, only the "cups" data were used in subsequent analyses. Arithmetic means rather than medians or geometric means were used because there was little skewness in the distributions.

Figure 1 shows the mean estimated equivalent number of cups of coffee as a function of different caffeine doses received. An analysis of variance showed that there was a significant main effect [$F(3,36) = 5.4, p < .004$]: larger doses received higher magnitude estimates (SDs were 1.1, 1.0, 1.8, and 2.3, for the four doses in ascending order). Statistical analysis confirmed that the linear trend was significant [$F(1,36) = 15.2, p < .001$], and no higher order trends reached significance.

Figure 2 shows highly similar results for category ratings. The ratings were analyzed by assuming that the categories were equidistant on a 5-point scale and using arithmetic means. Again the means for the various doses differed significantly [$F(3,36) = 2.9, p < .05$]: larger doses received higher category ratings (SDs were 1.0, 1.1, 1.4, and 1.5, for the four doses in ascending order). The linear trend was significant [$F(1,36) = 8.64, p < .01$], and no higher order trends reached significance ($F_s < 1$).

Discussion

These results indicate that untrained persons can, in fact, judge the dosage of caffeine they have consumed with some accuracy and in the absence of the usual gustatory/olfactory cues. A comparison of the pilot study with Experiment 1 indicates that there must be sufficient delay between the time of ingestion and the time when the judgment is made. The delay is necessary for the major portion of the pharmacological effects of caffeine to occur and may be important in the epidemiology of caffeinism. A person could come to expect immediate effects from caffeine because it is often consumed in a socially stimulating situation, where psychological sources of arousal provide immediate effects. When social stimulation is absent and pharmacological effects are not forthcoming, that person might take repeated dosages in an attempt to derive an immediate stimulating effect. Unfortunately, even heavy dosages do not provide immediate feedback and such binges may become a pattern that eventually leads to habitual heavy caffeine use. This may be the reason why the excessive tension and anxiety reported by heavy drinkers are not effective deterrents to the habit. The explanation given here places the roots of excessive behavior in a mistaken perceptual cycle, rather than, say, a need to counteract cortical inhibition (Eysenck, 1965). It is also important to consider polydrug patterns: caffeine may counteract the effects of nicotine or alcohol, and minor tranquilizers may counteract caffeine (Ayers, Ruff, & Templer, 1976; Greden, Procter, & Victor, 1981).

The agreement between magnitude estimates and category ratings implies that these two methods are essentially equivalent for this modality within these ranges of stimulation. The usual curvilinear relationship between

magnitude and category scales (Stevens & Galanter, 1957) does not appear, perhaps because the response range assigned by the experimenter's categories is no different from what the subject would choose in unconstrained magnitude estimates (cf. Foley, Cross, Foley, & Reeder, 1983) or because it is difficult to characterize functions precisely from only four data points. The usual range and frequency effects on scales (Parducci, 1974) are absent because only one stimulus was used. Thus, in the future, experimenters would be justified in using whichever method is more convenient.

EXPERIMENT 2

The second most frequently listed reason for caffeine consumption is its taste. Some sources report that coffee drinkers prefer the taste of caffeinated coffee to the taste of decaffeinated coffee (e.g., "Instant Coffees," 1979). Moreover, food manufacturers have used caffeine as a flavoring agent (*Food Chemicals Codex*, 1981). Nonetheless, the role of caffeine's taste in foods is largely unknown or unproven.

A number of studies have investigated the ability to detect caffeine flavor in isolation (e.g., McBride, 1983; Schutz & Pilgrim, 1957). Some studies have also investigated simple interactions between two flavors presented simultaneously—bitterness (caffeine) and saltiness, caffeine and monosodium glutamate, etc. (see the summary in Amerine, Pangborn, & Roessler, 1965). Such studies have not focused on caffeine itself, but rather have used it as an example of "bitter" less objectionable than quinine.

Three published reports concern caffeine in consumer products. Mackey and Valassi (1956) found that the detection threshold for caffeine was lower in a water solution than in tomato juice or custard. King (1937) found no correlation between judging pure solutions and rating samples of substances containing caffeine. These results highlight the problems in generalizing from pure solutions to more complex taste stimuli. The third study ("Caffeine," 1981) involved four days of training and a sensitive test (choosing the odd member from among three samples). Subjects had to discriminate between a noncaffeinated orange soda and a subtly fortified version of the same soda (an added 21 mg of caffeine per 6 oz; i.e., .00061 moles/liter). Although such a test situation may reveal the limits of discrimination abilities, such abilities may be of no practical importance in everyday tasting situations.

The main goal of Experiment 2 was to test whether *untrained* humans have the ability to sense graded gustatory and olfactory cues of caffeine in coffee. Although it was not possible (or desirable) to eliminate all laboratory constraints, in Experiment 2 we avoided repeated trials, progressive training, side-by-side comparisons, and isolation from competing stimuli. In order to provide a comparison between caffeine perception via taste and perception via bodily effects, we assumed that a dose received in cap-

sules in Experiment 1 should be paired with the same dose in exactly one cup of coffee in Experiment 2. Given 0-600 mg per cup, several of these beverage concentrations were higher than those encountered in everyday experience. If a functional relationship emerged with the large concentration differences used in the present experiment, then fine-grained acuity could be tested in future experiments with much smaller differences. Because habitual heavy use of caffeine would be expected to affect sensitivity to caffeine, Experiment 2 tested for differences between heavy users and others (moderate-, light- and nonusers). Finally, Experiment 2 was designed to test for a possible difference between males and females. [Two previous studies found sex differences, but in opposite directions ("Caffeine," 1981; King, 1937)].

Method

Subjects. The 89 subjects (51 females and 38 males) were college students from the same population as those in Experiment 1. They were recruited and tested in lobbies of two campus buildings, and participation was gratis.

Stimuli. For each testing session, four containers of coffee were prepared in the following manner. In each of four thermal serving pots, four teaspoons (approximately 9.2 g) of Hills Brothers decaffeinated instant coffee were mixed with 24 oz of boiling water. Three preweighed amounts of reagent caffeine were added to three of the coffee mixtures to produce concentrations of 200, 400, and 600 mg per 6-oz cup; one thermos had no added caffeine. Because decaffeinated coffee has residual amounts of caffeine close to 2 mg per 6-oz cup, and the reagent caffeine was weighed on a pan balance to an accuracy of ± 10 mg, it was estimated that the actual concentrations were within 5% of .00006, .006, .012, and .017 moles/liter. The coffee was served at comfortable drinking temperature. Tactile and visual cues were freely available since such information does not indicate concentration level.

Response scale. Observers were asked to "rate the caffeine taste" of their sample on a category scale similar to that of Experiment 1: (0) "decaffeinated coffee," (1) "weaker than normal caffeinated coffee," (2) equivalent to normal caffeinated coffee," (3) "stronger than normal caffeinated coffee," (4) "much stronger than normal caffeinated coffee." No magnitude estimates were obtained.

Procedure. Students who walked by the "taste test" were invited to participate. Those who volunteered signed a consent form, were informed that the coffee samples had various amounts of caffeine (including no caffeine), filled out the caffeine habits questionnaire, were poured a 1-oz (30-ml) sample from one of the four coded thermos pots, tasted the sample, and marked their judgment on the 5-point scale.

Results

Observers were able to evaluate accurately the concentration of caffeine in the samples. Figure 3 shows the mean category ratings based upon taste. With higher levels of concentration, observers gave higher ratings on the category scale.

A three-way analysis of variance was performed on the factors caffeine level \times sex \times caffeine habits. The simple main effect of caffeine level reached significance [$F(3,73) = 3.95, p < .02$]. A trend analysis performed on caffeine levels showed a significant positive linear trend [$F(1,73) = 11.82, p < .01$]; no higher order trends reached significance. The factor of sex did not reach significance ($F < 1$), although women's ratings (Mean = 2.25) were slightly higher than men's ratings (Mean =

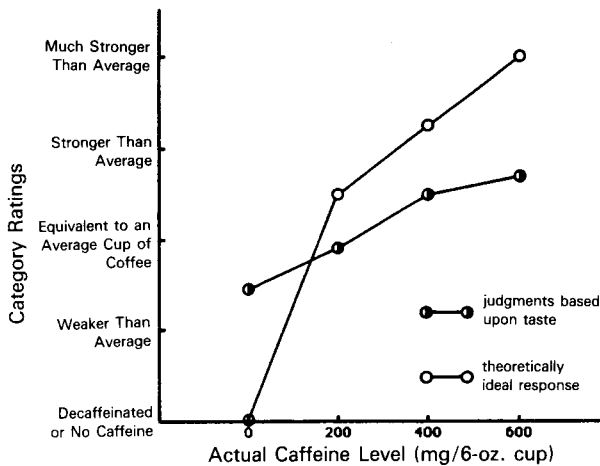


Figure 3. Mean category ratings of caffeine level based upon ad-lib sips of a sample of coffee with judgments made immediately, Experiment 2.

1.97). There were two levels of the caffeine habits factor: (1) heavy users (habitual consumption of 400 mg or more of caffeine per day) and (2) others. Caffeine habits did not reach significance. None of the two-way interactions reached significance, nor did the three-way interaction (level \times sex \times habits).

Discussion

The results of Experiment 2 show that observers can rate taste on a category scale to reflect accurately levels of caffeine concentration. This implies that caffeine in the concentrations used here imparts a noticeable taste to coffee, a heavily flavored carrying fluid. Thus, it is somewhat puzzling that subjects judged decaffeinated coffee as having only slightly less caffeine than "normal" caffeinated coffee. This unexpected finding may be due to the fact that Weber's law does not hold at these low stimulus levels. On the other hand, the subjects may have been unable to distinguish bitterness due to caffeine from the various other bitter flavors in decaffeinated coffee.

Without taste training, the implied training in repeated trials, or the chance to make side-by-side comparisons of samples, the subjects in this experiment faced what one would expect to be a very difficult task. Nonetheless, the perceptual abilities demonstrated here corroborate and extend previous studies that used specialized laboratory conditions and/or indirect scaling techniques. For example, when the 5-point scale from this experiment is mapped onto McBride's (1983) 13-point scale, our means fall within two of his standard errors of his means.

The results of Experiments 1 and 2 taken together (Figures 2 and 3) show responses to two sources of information about levels of caffeine, namely, gustatory/olfactory cues and pharmacological effects. Both sets of responses are roughly linear against caffeine level, with each 200 mg increase producing roughly a one-category change in rating. Both show attenuation of slope relative to the "ideal response" in spite of precautions to avoid

biases that produce such attenuation. Because gustatory cues are immediately available, they appear to be a valuable supplement to sensed pharmacological effects for people who are monitoring their own states. The difference in absolute response level (Mean = 1.70 in Figure 2 and Mean = 2.13 in Figure 3) suggests one of two things: Taste is more salient than bodily effects, or taste judgments were more synchronized with maximum taste effects than arousal judgments were with maximum bodily effects. (It is likely that across subjects, the rise and fall patterns vary more for bodily effects than for taste effects.)

It has often been suggested that taste plays a role in regulating the intake of substances (including caffeine) through a simple hedonic mechanism of approach and avoidance. Chances of survival would be improved if more toxic substances were more detectable; however, the evidence is that there is little relationship between toxicity and detectability (Amerine et al., 1965). Perhaps avoidance of all bitter substances is too simple a principle. An organism could be more adaptable and have improved chances of survival through a graded response to concentrations coupled with the ability to learn which substances were toxic. All high concentrations would be avoided until they were learned to be nontoxic.

Experiment 2 suggests the need for greater caution regarding the use of caffeine placebos in attempts to separate psychological and pharmacological effects. If psychological expectations are related only to the presence or absence of caffeine, then decaffeinated coffee will be effective in inducing the belief that some caffeine was received. On the other hand, if the size and kind of expectations are related to the amount of caffeine, then the surprising accuracy of subjects' scaling suggests that perceptual cues will override an experimenter's deceptive claims regarding the strength of the placebo.

Unlike previous experiments ("Caffeine," 1981; King, 1937), the present experiment did not find sex differences in the ability to taste caffeine. It is unclear whether such discrepancies are due to differences among experiments in designs, carrier substances, caffeine levels, etc.

The present experiment used relatively large differences in caffeine levels, on the order of magnitude of the difference between a tumbler of water and a strongly brewed mug of drip coffee. There remains the question of whether smaller differences could be scaled as accurately, or even detected at all.

EXPERIMENT 3

The third experiment was designed to compare the relative salience of (low dose) caffeine-specific flavor characteristics with that of brand-specific flavor characteristics. Rather than use the high caffeine concentration of brewed coffee, we chose to use the caffeine level of a cup of instant coffee. Published work (e.g., "Caffeine," 1981) suggested that this task would be difficult; therefore, the design involved practice with repeated trials and included conditions with and without feedback. In addition, this experiment again tested for sex differences.

Experiment 3 was patterned after studies of the identification of consumer beverages on the basis of taste. The methodological inadequacies of many such studies were avoided by incorporating a number of improvements. (1) Subjects were informed of the range of possibilities in order to restrict irrelevant guessing (cf. Thumin, 1962). (2) Stimuli were presented in pairs, a more sensitive procedure than presenting all stimuli simultaneously (cf. Thumin, 1962). Paired *presentations* were also used, because it is quite difficult to identify an isolated sample without another sample for reference. This method is similar to the standard technique of paired *comparisons*, except that two separate responses are required rather than one response regarding the relationship between the two samples in a pair. (3) Subjects were permitted to make the same response for both members of a pair, because requiring different responses might bias judgments and propagate errors (cf. Lane, Zychowski, & Lelii, 1975).

Method

Subjects. The 63 volunteers (30 males and 33 females) were recruited in the same manner as in Experiment 1. Subjects received extra course credit for their participation.

Stimuli. Four beverages were used—two brands of instant coffee (Taster's Choice and Hills Brothers), each presented in both its caffeinated and decaffeinated forms. These brands were chosen for their availability in both forms in local stores, and for their distinctive flavors [Taster's Choice has a bean character, Hills Brothers has a grain character ("Instant Coffees," 1979)]. The caffeinated versions have essentially identical caffeine content, 57 mg per 6-oz cup for Taster's Choice versus 59 mg for Hills Brothers ("Instant Coffees," 1979). The coffee was kept in insulated coffee pots from which 1-oz samples were served.

Design. There were 12 trials with a pair of samples on each trial, for a total of 24 samples, that is, six judgments of each of the four beverages. Each type of coffee was paired twice with every other type so that no trial contained a pair of identical samples. The order in which the two samples in a pair were presented was reversed the second time that pair was presented. The same random order of trials was used for all subjects. Half received feedback at the end of each trial, while the other half did not.

Procedure. Subjects were tested individually after completing the caffeine habits questionnaire. To eliminate visual cues, they were blindfolded while tasting. The subjects' tasks were to indicate for each sample (1) whether it tasted like Brand A or Brand B and (2) whether it was caffeinated or decaffeinated. Before the trials began, subjects tasted the four coffees and were told the proper labels. On each trial of the experiment proper, subjects tasted and then attempted to identify each sample from the nonidentical pair. Between trials, subjects ate a piece of cracker and took a sip of water. Following each trial, subjects in the feedback condition were informed of the correct brands and caffeinations. The 12 trials required about 10 minutes.

Results

Scores were number of correct identifications (of brand and of caffeination) for each task across 12 trials. A preliminary two-way analysis of variance showed no differences between males and females and no interaction between sex and task ($p < .26$). Hence, scores for males and females were pooled for all further analyses.

The data were then subdivided in a different way. For each subject, the number of correct identifications was tallied separately for the first and last half of the trials.

Table 1
Mean Number of Correct Identifications for
Each Condition in Experiment 3

Brands Task		Caffeination Task	
Without Feedback	With Feedback	Without Feedback	With Feedback
First 6 Trials			
6.52	6.44	5.74	6.38
Last 6 Trials			
6.68	6.91	5.35	6.72

Note—Chance = 6.00, all data points based on 31 or 32 subjects.

This trials factor was combined with feedback and task. The means of these scores over subjects are given in Table 1. Most entries are close to chance. The two highest scores (6.91 and 6.72) are in conditions with feedback and after practice. The lowest score is below chance and is for identification of caffeine without feedback on the last 6 trials.

A three-way analysis of variance showed that only the simple main effect of task was significant [$F(1,61) = 3.998$, $p < .05$]. The other main effects, the two-way interactions, and the three-way interaction were not significant. It is worth noting, however, that the main effect of feedback approached significance [$F(1,61) = 3.13$, $p < .10$].

Scores were pooled across the feedback conditions and were added across trials in order to analyze the task factor further. Caffeination identification was at chance level: 12.11 correct from 24 samples, whereas 12 would be expected by chance. On the other hand, brand identifications were better than chance (Mean = 13.27, $Z = 4.12$, $p < .00003$), although performance was only 11% better than mere guessing.

Discussion

Although caffeine affects flavor in concentrations of the level of brewed and drip coffees (200 mg per 6-oz cup, Experiment 2), it has a negligible role in concentrations of the level of instant coffee (58 mg per 6-oz cup). Low concentrations may be discriminable after training, but do not seem to be easily discriminable in everyday situations. Individuals who need to control their caffeine intake will need perceptual training or will have to rely on nonperceptual cues. If these results generalize to tea, soft drinks, and other products with low caffeine concentrations, then caffeine can be said to be a relatively minor flavor component.

The two brands tested in Experiment 3 use two different decaffeination processes (water spray and methylene chloride, "What's in Your Decaf Coffee?," 1982). This difference may have exaggerated the usual discriminability of brands, which is based only upon combinations of coffee beans, packaging, freshness, etc. This distracting difference may account for the progressively lower than chance scores on judging caffeination without feedback: subjects may have perseverated on one process as decaffeinated and miscalled the other process caffeinated.

Water and crackers eliminate carryover effects in soft-drink tasting, in which smell plays a minor role. However, water and crackers may not be sufficient in coffee tasting, in which aroma plays a larger role. In future experiments it would be advisable to test whether olfactory purging between trials improves performance.

GENERAL DISCUSSION

Hall, Bartoshuk, Cain, and Stevens (1975) have found that taste thresholds for caffeine correlate with PTC (phenylthiourea or phenylthiocarbamide) thresholds and show bimodal distributions. The less sensitive or "taste blind" individuals also show lower perceived bitterness for suprathreshold concentrations of caffeine. Genetic studies attribute PTC differences in individuals to simple Mendelian dominance systems. It would be interesting to see if genetic status predicts caffeine sensitivity in any or all of the situations of Experiments 1, 2, and 3.

In summary, under one set of conditions (Experiments 1 and 2), subjects were able to scale caffeine concentrations accurately on the basis of gustatory/olfactory cues alone or on the basis of physiological effects alone. In part, these conditions resembled those that would be found in real-life situations: Only one judgment was made; there were no immediate standards for comparison; and bodily levels of caffeine were on the order of zero to several cups of coffee. Other aspects of these conditions were quite distinct from everyday life: there was a caffeine beverage concentration three or four times as strong as regular coffee; the procedure for administering caffeine led to precipitous rises in blood caffeine levels; and judgments were made near the peak of the relevant effects.

On the other hand, under a second set of conditions (Experiment 3), subjects were not able to identify whether instant coffee was caffeinated or decaffeinated. Most notably, these conditions involved low concentrations of caffeine. They also involved practice, feedback, and possibly distracting differences in brand flavors. The qualitative differences of brands, although subtle, are more salient than quantitative differences in caffeination in instant coffees. The untrained observer's ability to perceive caffeine taste in coffee is lost somewhere between 200 mg and 50 mg per cup. It should be pointed out that the perceptibility of low concentrations of caffeine on the basis of bodily effects has not yet been investigated.

REFERENCES

- AMERINE, M. A., PANGBORN, R. M., & ROESSLER, E. B. (1965). *Principles of sensory evaluation of food*. New York: Academic Press.
- AXELROD, J., & REICHTHAL, J. (1953). The fate of caffeine in man and a method for its estimation in biological material. *Journal of Pharmacology and Experimental Therapeutics*, **107**, 519-523.
- AYERS, J., RUFF, C. F., & TEMPLER, D. I. (1976). Alcoholism, cigarette smoking, coffee drinking and extraversion. *Journal of Studies on Alcohol*, **37**, 983-985.
- BENDIG, A. W., & HUGHES, J. B. (1953). Effect of amount of verbal anchoring and number of rating-scale categories upon transmitted information. *Journal of Experimental Psychology*, **46**, 87-90.
- BLOUNT, J. P., & COX, W. M. (1982, August). *State-dependent learning with caffeine in a college classroom setting*. Paper presented at the Ninetieth Annual Convention of the American Psychological Association, Washington, DC.
- BLOUNT, J. P., & COX, W. M. (1983, November). *Caffeine habits, beliefs, and cognitive effects among college students*. Paper presented at the Twenty-Fourth Annual Meeting of The Psychonomic Society, San Diego, CA.
- BRODY, J. E. (1982, April 21). Weaning the body from dependence on caffeine. *The New York Times*, p. 18.
- Caffeine: How to consume less (1981, October). *Consumer Reports*, pp. 597-599.
- COHEN, S., & BOOTH, G. H. (1975). Gastric acid secretion and lower-esophageal-sphincter pressure in response to coffee and caffeine. *New England Journal of Medicine*, **293**, 897-899.
- COLTON, T., GOSSELIN, R. E., & SMITH, R. P. (1968). The tolerance of coffee drinkers to caffeine. *Clinical Pharmacology and Therapeutics*, **9**, 31-39.
- COX, W. M., & BLOUNT, J. P. (1984, August). *Recent psychological research on caffeine*. Symposium at the 92nd Annual Convention of the American Psychological Association, Toronto.
- DAWBEN, T. R., KANNEL, W. B., & GORDON, T. (1974). Coffee and cardiovascular disease. *New England Journal of Medicine*, **291**, 871-874.
- EYSENCK, H. J. (1965). *Smoking, health, and personality*. New York: Basic Books.
- FOLEY, H. J., CROSS, D. V., FOLEY, M. A., & REEDER, R. (1983). Stimulus range, number of categories, and the "virtual" exponent. *Perception & Psychophysics*, **34**, 505-512.
- Food Chemicals Codex* (3rd ed.). (1981). Washington, DC: National Academy Press.
- GILBERT, R. M. (1976). Caffeine as a drug of abuse. In R. J. Gibbins, Y. Israel, H. Kalant, R. E. Popham, W. Schmidt, & R. G. Smart (Eds.), *Research advances in alcohol and drug problems*. New York: Wiley.
- GOLDSTEIN, A., & KAIZER, S. (1969). Psychotropic effects of caffeine in man. III. A questionnaire survey of coffee drinking and its effects in a group of housewives. *Clinical Pharmacology and Therapeutics*, **10**, 477-488.
- GREDDEN, J. F., PROCTER, A., & VICTOR, B. (1981). Caffeinism associated with greater use of other psychotropic agents. *Comprehensive Psychiatry*, **22**, 565-571.
- HALL, M. J., BARTOSHUK, L. M., CAIN, W. S., & STEVENS, J. C. (1975). PTC taste blindness and the taste of caffeine. *Nature*, **253**, 442-443.
- Instant Coffees. (1979, October). *Consumer Reports*, pp. 569-572.
- JESTEADT, W., WIER, C. C., & GREEN, D. M. (1977). Intensity discrimination as a function of frequency and sensation level. *Journal of the Acoustical Society of America*, **61**, 169-177.
- JICK, H., MIETTINEN, O. S., NEFF, R. K., SHAPIRO, S., HEINONEN, O. P., & SLONE, D. (1973). Coffee and myocardial infarction. *New England Journal of Medicine*, **289**, 63-67.
- KING, F. B., (1937). Obtaining a panel for judging flavor in foods. *Food Research*, **2**, 207-219.
- LANE, S. H., ZYCHOWSKI, J., & LELI, K. (1975). Cola and diet cola identification and level of cola consumption. *Journal of Applied Psychology*, **60**, 278-279.
- MACKEY, A. O. & VALASSI, K. (1956). The discernment of primary tastes in the presence of different food textures. *Food Technology*, **10**, 238-240.
- MANBER, M. (1976). The medical effects of coffee. *Medical World News*, **17**(2), 63-73.
- MCBRIDE, R. L. (1983). A JND-scale/category-scale convergence in taste. *Perception & Psychophysics*, **34**, 77-83.
- PARDUCCI, A. (1974). Contextual effects: A range-frequency analysis. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 2.) New York: Academic Press.
- POULTON, E. C. (1979). Models for biases in judging sensory magnitude. *Psychological Bulletin*, **86**, 777-803.
- REVELLE, W., HUMPHREYS, M. S., SIMON, L., & GILLILAND, K. (1980). The interactive effect of personality, time of day, and caffeine: A test of the arousal model. *Journal of Experimental Psychology: General*, **109**, 1-31.

- RITCHIE, J. M. (1975). Central nervous system stimulants. In L. S. Goodman & A. Gilman (Eds.), *The pharmacological basis of therapeutics* (5th ed., pp. 367-378). New York: MacMillan.
- SCHUTZ, H. G. & PILGRIM, F. J. (1957). Differential sensitivity in gustation. *Journal of Experimental Psychology*, **54**, 41-48.
- STEVENS, S. S. & GALANTER, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, **54**, 377-411.
- TEGHTSOONIAN, M., BECKER, E., & EDELMAN, B. (1981). A psychophysical analysis of perceived satiety: Its relation to consumatory behavior and degree of overweight. *Appetite: Journal for Intake Research*, **2**, 217-229.
- THUMIN, F. J. (1962). Identification of cola beverages. *Journal of Applied Psychology*, **46**, 358-360.
- What's in your decaf coffee? (1982, November). *Nutrition Action*, p. 3.

(Manuscript received February 6, 1985;
revision accepted for publication June 20, 1985.)