Preference can be more powerful than detection of oddity as a test of discriminability

A. W. MACRAE and E. N. GEELHOED
University of Birmingham, Birmingham, England

Subjects presented with sets of three samples, two of distilled water and one of tap water, were significantly more consistent in choosing the tap water as preferable than they were in identifying it as the odd sample in the set. The result is opposite to the prediction of high-threshold models of sensory discrimination, which say that if a difference is not noticed, preferences will be random, whereas if a difference is noticed, preferences may still be in either direction. The result can be quantitatively explained by a model advanced by Frijters to explain an analogous anomaly found with the triangle test used in the food industry. Applying his model to the observed proportions yields essentially equivalent estimates of sensory difference ($d' = 1.5$, approximately) from the two tasks, and a direction of preference almost unanimously in favor of the tap water that was used. Since the model predicts that the proportion of subjects choosing the odd item will depart further from chance in the preference task than in the oddity task, the former has greater power to reject the null hypothesis of no sensory difference if one exists and if preference is overwhelmingly in one direction.

The food industry makes considerable use of the *triangle test*, in which a subject is presented with three samples, two of them identical and one different. The task is to identify the different sample. Although this task has a superficial similarity to three-alternative forced choice (3-AFC), there are important differences between them that Frijters (1979b) has discussed in detail. The task's principal virtue is that it does not require the subject to identify the attribute responsible for the difference between the samples, which makes it particularly suitable for use with complex, multiattributed stimuli such as foods. Thus it is regarded as a pure test of discriminability: if *any* difference is detectable, it can be used to identify the odd member of the triangle.

The *two-stage triangle test* elaborates on the basic design. After the triangle test has been administered, a subject is asked to characterize the difference between the "odd" stimulus and the other two by means of a sensory reevaluation of the same triangle. But the conventional wisdom of triangular testing holds that the two-stage test has adverse effects on discrimination and is therefore undesirable.

At a University Open Day, we demonstrated two of the basic procedures used in food evaluation: the classical triangle test, and the classical pair-comparison test. Both tests were informally, but quite carefully, conducted with

more than 130 visitors; we used samples of distilled water and tap water, because one of our aims was to learn more about the sensory properties of our local tap water in connection with other research in progress.

We expected the triangle test (asking for identification of the odd sample) to show that the two types of water were discriminable but expected the pair-comparison test (asking which of two samples was preferred) to have results closer to chance, because, of those who were sensitive to the differences between the samples, some would prefer tap and some prefer distilled water. The results we obtained were the opposite. We found a preference for tap water in the pair-comparison test that was much stronger ($\chi^2 = 34.0$, $C = 0.45$) than the detectability of the difference in the triangle test ($\chi^2 = 3.9$, $C = 0.17$).

Our conclusion that the results were opposite to expectation was supported by various analyses of the data, but it might not be considered compelling, because the informality of the procedure meant that not quite all of the subjects had performed both tasks and because one task required the subjects to judge three samples, whereas the other required them to judge only two. We therefore conducted a controlled experiment to avoid these objections.

METHOD

Subjects

The subjects were 100 volunteers, most of whom were staff and students in the School of Psychology at the University of Birmingham.

Materials

Tap water was obtained from the same tap in the School of Psychology throughout the experiment. Distilled water was prepared in an automatic still. Before each testing day, between 1600 and 1700, fresh samples of each water were obtained and stored over-

This work was supported by a research contract from the Ministry of Agriculture, Fisheries and Food. The results of the research are the property of the Ministry and are Crown copyright. We are grateful to D. M. Ennis, J. E. R. Frijters, G. R. Lockhead, and N. Macmillan for valuable comments on earlier drafts. Correspondence should be addressed to A. W. MacRae, School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, England.

night in screw-topped 2-l glass bottles to stabilize them and equalize their temperatures in a room where an average temperature of $16^{\circ} \pm 2^{\circ} \text{C}$ was maintained.

White, expanded polystyrene thermal cups were used to present the samples. Each cup was labeled with a three-digit random number, generated in such a way that every number consisted of three different digits and the last digit was never 0 or 5. The water was served at a temperature of $15^{\circ} \pm 2^{\circ} \text{C}$ equalized to better than 0.2° within each set. Each sample was of about 60 ml.

Procedure

Subjects were recruited and tested individually, using a mobile booth in various parts of the building. There were two tasks, in each of which three cups were presented, two containing distilled water and one containing tap water. The cups were presented in a row on a tray, with the cup containing the tap water equally often on the left, right, and center. Each subject performed both tasks, but only once and in a single session of about 3 min. The order of the tasks was alternated for successive subjects.

In each task, subjects were told to taste the three samples in any order and to taste as many times as required. They were told that it was safe to swallow a sample, but that should they wish to, they could spit it out into a bucket that was provided. In Task 1, they were told that two cups contained samples from the same source and one that was different, and they were asked to determine which sample was different from the other two. Task 2 used a different set of three samples, and it was not stated that two samples were from the same source (though they were). The subjects were asked to say which of the three samples they liked best.

RESULTS

Table 1 summarizes the results. The columns show the outcomes when the question was, "Which sample is the odd one?" The tap water was always the odd one, so the column headed Tap signifies the number of correct responses to the oddity task. The column Distilled contains the numbers who chose one of the distilled water samples as the odd one and therefore made the wrong response. From a total of 100 subjects, 53 correctly selected the odd one. The rows show responses to the question, "Which sample do you like best?" These preference data show that 72 of the 100 subjects preferred the single sample of tap water over the two samples of distilled water.

Chance performance in each task would be 33.3 selections of tap water. Both outcomes exceed that, but the preference results do so by a greater margin, and the difference between the proportions selecting tap water in the two tasks is highly significant ($\chi^2 = 8.44$, $p < .005$).

Of the 53 who correctly identified tap water in one set as the odd sample, about 80% gave tap water as their

preference in the other set that they tasted. More surprisingly, perhaps, is that of the 47 who gave the wrong response in the oddity task and who, according to a high-threshold theory, would be considered nondiscriminators, more than 63% (rather than the chance proportion of 33.3%) gave tap water as their preference. The departure from chance is highly significant ($\chi^2 = 19.6$, $p < .001$). Clearly, a dichotomization into discriminators and nondiscriminators will not do. Those who chose distilled water as preferred were essentially random in locating the odd item (39% rather than 33.3%, $\chi^2 < 1$).

DISCUSSION

Something surprising is going on. Just over half of the subjects selected tap water as the odd one, whereas almost three quarters expressed a preference for tap water—but they could not have done that unless there was a difference between tap and distilled water that they were able to detect. It appears that a substantial number of "nondiscriminators" must have discriminated when the task was to express a preference.

Factors such as the faint but detectable color difference between the two waters cannot explain the result because, without exception, any real differences that made preferences more pronounced should also have helped the oddity task. One contributing factor may have been that preferences are neither right nor wrong. Failure was impossible and that may have put subjects at ease and allowed them to discriminate better. It is also possible that the biological advantage of making good decisions about pleasant and unpleasant options results in the evolution of efficient processing mechanisms for hedonic variables.

We cannot discount these possibilities, but we believe that the effect is in fact related to one that has been called the paradox of the discriminatory nondiscriminators (Byer & Abrams, 1953; Gridgeman, 1970). In the two-stage triangle test, a significant number of subjects who were unable to identify the odd stimulus in a triangle test gave nonrandom answers when they were asked to characterize the difference between the samples.

Frijters (1979a) resolved the anomaly by invoking a model whereby the triangle and the 3-AFC require different decision strategies. For the 3-AFC, the optimal decision rule is to select the most intense stimulus. For the triangle test, the optimal decision rule is to choose the smallest difference between pairs of stimuli and identify as the odd item the one not in the pair. These decision rules yield the same probability of success (1/3) when there is no sensory difference between samples (that is, when $d' = 0$), but otherwise their success rates are different. In a 3-AFC task, a larger proportion of correct responses is predicted for any particular size of d' . Frijters, Kooistra, and Vereijken (1980) give tables relating triangle-test and 3-AFC performance to d' , for both normal and logistic distributions of sensory effect.

These models can be applied to our oddity task (decisions about pair differences) and our preference task (decisions about individual items). They predict the results

Table 1
Numbers of Subjects Selecting Each Type of Water as the Odd Sample (Columns) and as Their Preferred Sample (Rows)

Preferences	Oddity Judgments		Totals
	Tap	Distilled	
Tap	42	30	72
Distilled	11	17	28
Totals	53	47	100

Note—The row and column data were obtained from the same 100 subjects but from two different sets of samples. In every trial, one tap water and two distilled water samples were presented.

found here, provided that preference is almost uniformly in one direction. It is as though the preference task involved deciding, by means of 3-AFC, which item had most of a known attribute of "pleasantness," whereas the oddity task was a triangle test with an unknown attribute responsible for "difference," thus requiring the subject to evaluate the relative sizes of the three differences between pairs in the triangle. The values of d' derived by Frijters et al. for our observed proportions of 0.53 (triangle) and 0.72 (3-AFC) are 1.62 and 1.31, respectively, for a normal distribution, or 1.51 and 1.25 for a logistic. The preference d' is the smaller of the two, perhaps indicating that the hedonic difference between tap and distilled water is not in the same direction for everyone, with a few people disliking the attributes that lead most people to prefer the tap water. However, the confidence intervals of our observed proportions are such that when converted to equivalent d' 's in that way, the d' for each task lies inside the confidence bounds of the other, so the difference should probably not be interpreted. Both tasks thus give much the same estimate of d' , but the preference task leads to a larger departure from chance in the proportion of subjects selecting the odd item and thus

offers greater statistical power. That advantage cannot be guaranteed in general, however, because it will not occur unless almost all subjects have the same preference order for the materials used. In the present case, almost everyone must have preferred our tap water.

REFERENCES

- BYER, A. J., & ABRAMS, D. (1953). A comparison of the triangular and two-sample taste-test methods. *Food Technology*, **7**, 185-187.
- FRIJTERS, J. E. R. (1979a). The paradox of discriminatory non-discriminators resolved. *Chemical Senses & Flavour*, **4**, 355-358.
- FRIJTERS, J. E. R. (1979b). Variations of the triangular method and the relationship of its unidimensional probabilistic models to three-alternative forced-choice signal detection theory models. *British Journal of Mathematical & Statistical Psychology*, **32**, 229-241.
- FRIJTERS, J. E. R., KOOISTRA, A., & VERELUKEN, P. F. G. (1980). Tables of d' for the triangular method and the 3-AFC signal detection procedure. *Perception & Psychophysics*, **27**, 176-178.
- GRIDGEMAN, N. T. (1970). A reexamination of the two-stage triangle test for perception of sensory differences. *Journal of Food Science*, **35**, 87-91.

(Manuscript received September 14, 1990;
revision accepted for publication September 19, 1991.)