not want the consumer to perceive a difference in the product. A properly executed discrimination test with sufficient power indicating that the two ice cream formulations are not perceptibly different would allow the company to make the substitution with lowered risk. This is an ideal use of sensory discrimination testing. Discrimination testing may also be used when a processing change is made which the processor hopes would not affect the sensory characteristics of the product. In both of these cases the objective of the discrimination test is not to reject the null hypothesis, this is also known as similarity testing.

However, when a company reformulates a product to make a "new, improved" version then the discrimination test could be used to indicate that the two formulations are perceived to be different. In this case the objective of the discrimination is to reject the null hypothesis. If the data indicate that the two formulations are perceptibly different then the sensory scientist has to do a test that would indicate that the "new" formulation is perceived to be an improvement by the targeted consumer (see Chapters 13–15).

If the difference between the samples is very large and thus obvious, discrimination tests are not useful. If preliminary bench testing indicates that the two samples will be perceptibly different to all panelists then these discrimination tests should not be used. In such cases it may be useful to do scaling techniques to indicate the exact magnitude of the difference between the samples (see Chapter 7). In other words, discrimination testing is most useful when the differences between the samples are subtle. However, these subtle differences make the risk of Type II errors more likely (see later in this chapter and Appendix E).

Discrimination tests are usually performed when there are only two samples. It is possible to do multiple difference tests to compare more than two products but this is not efficient or statistically defensible. Usually ranking or scaling techniques will prove to be more effective (see Chapter 7).

There are a number of different discrimination tests available including triangle tests, duo-trio tests, paired comparison tests, *n*-alternative forced choice tests, tetrad tests (Frijters, 1984), polygonal and polyhedral tests (Basker, 1980). In Chapter 1, we briefly outlined the history associated with the triangle, duo–trio, and paired comparison tests. In the following section the more usual discrimination tests and their uses are described in more detail.

4.2 Types of Discrimination Tests

See Table 4.1 for a summary of the types of available discrimination tests and Table 4.2 for the process of doing a discrimination test.

4.2.1 Paired Comparison Tests

There are two analytical sensory forms of this test, namely the directional paired comparison (also known as the two-alternative forced choice) test and the difference paired comparison (also known as the simple difference or the same/different) test. The decision to use one or the other form is dependent on the objective of the study. If the sensory scientist knows that the two samples differ only in a specific sensory attribute then the two-alternative forced choice (2-AFC) method is used. In fact, as we will discuss in Chapter 5, it is always more efficient and powerful to use a directional paired comparison test specifying the sensory attribute in which the samples differ (if known) than to ask the panelists to identify the different sample. On the other hand, if the sensory scientist does not know in which sensory attribute(s) the samples differ than other techniques, such as the difference paired comparison must be employed, despite the subsequent loss of power.

For both paired comparison methods the probability of selection of a specific product, by chance alone (guessing), is one chance in two. However, as explained in Chapter 5 the situation is a little fuzzier for the same/different test where the probability is affected by the individual panelist's decision criterion. In both cases the null hypothesis states that in the long run (across all possible replications and samples of people) when the underlying population cannot discriminate between the samples they will pick each product an equal number of times. Thus the probability of the null hypothesis is $P_{\text{pc}} = 0.5$. Remember that P_{pc}, the proportion that we are making an inference (a conclusion) about, refers to the proportion we would see correct in the underlying population (and not the proportion correct in our sample or the data). That is why statistical hypothesis testing is part of inferential statistics. What the null hypothesis states in mathematical terms can also be verbalized as follows: If the

Table 4.1 Types of available discrimination tests

Class of test	Test	Samples: inspection phase	Samples: test phase	Task/instructions	Chance probability
Oddity	Triangle	(None)	A, A′, B (or A, B, B′)	Choose the most different sample	1/3
Matching	Constant reference duo–trio	Ref-A	A, B	Match sample to reference	1/2
	Balanced reference duo–trio	Ref-A, Ref-B	A, B	Match sample to reference	1/2
	ABX	Ref-A, Ref-B	A (or B)	Match sample to reference	1/2
	Dual standard	Ref-A, Ref-B	A, B	Match both pairs	1/2
Forced choice	Paired comparison	(None)	A, B	Choose sample with most of specified attribute	1/2
	3-AFC	(None)	A, A′,B	(Same)	1/3
	n-AFC	(None)	$A_1 - A_{n-1}$, B	(Same)	1/*n*
	Dual pair	(None)	A, B and A, A′	Choose A, B (different pair)	1/2
Sorting	Two out of five	(None)	A, A′, B, B′, B″	Sort into two groups	1/10
	4/8 "Harris–Kalmus"	(None)	$A_1 - A_4$, $B_1 - B_4$	Sort into two groups	1/70
Yes/no	Same–different	(None)	Pairs: A, A′ or A, B	Choose response: "Same" or "different"	N/A[a]
(Response choice)	A, not-A	Ref-A	A or B	Choose response: "A" or "not-A"	N/A[a]

[a]For the yes/no tests, a criterion may be set by each individual and therefore the probability may not be equal to 1/2. See Chapter 5 for further discussion of criterion in yes/no tasks

Table 4.2 Steps in conducting a difference test

1. Obtain samples and confirm test purpose, details, timetable, and panelists' training (e.g., training with the process) with client.
2. Decide testing conditions (sample size, volume, temperature, etc.) and clear with client.
3. Write instructions to the panelists and construct ballot.
4. Recruit potential panelists.
5. Screen panelists for acuity.
6. Train to do specific difference test (can use colors or shapes or spiked samples).
7. Set up counterbalanced orders.
8. Assign random three-digit codes and label sample cups/plates.
9. Conduct test.
10. Analyze results.
11. Communicate results to client or end user.

underlying population cannot discriminate between the samples then the probability of choosing sample A (that is the P_A) is equal to the probability of choosing sample B (P_B). Mathematically, this may be written as

$$H_0 : P_A = P_B = \frac{1}{2} \quad (4.1)$$

However, as we will see the verbal forms of the alternate hypotheses for the two paired comparison tests differ.

4.2.1.1 Directional Paired Comparison Method (or the Two-Alternative Forced-Choice Method)

In this case, the experimenter wants to determine whether the two samples differ in a specified dimension, such as sweetness, yellowness, crispness. The two samples are presented to the panelist simultaneously and the panelist is asked to identify the sample that is higher in the specified sensory attribute. Figure 4.1 shows a sample score sheet. The panelist must clearly understand what the sensory specialist

Fig. 4.1 Example of a directional paired comparison (2-AFC) score sheet.

Date ______________
Name ______________

Please rinse your mouth with water before starting. There are two samples in each of the two paired comparison sets for you to evaluate. Taste each of the coded samples in the set in the sequence presented, from left to right, beginning with Set 1. Take the entire sample in your mouth. NO RETASTING. Within each pair, circle the number of the sweeter sample. Rinse with water between samples and expectorate all samples and water. Then proceed to the next set and repeat the tasting sequence.

Set		
1	________	________
2	________	________

means by the specified dimension and the panelist should therefore be trained to identify the specified sensory attribute. The panelist should also be trained to perform the task as described by the score sheet. The directional paired comparison test has two possible serving sequences (AB, BA). These sequences should be randomized across panelists with an equal number of panelists receiving either sample A or sample B first.

The test is one tailed since the experimenter knows which sample is supposed to be higher in the specified dimension. The alternative hypothesis for the directional paired comparison test is that if the underlying population can discriminate between the samples based on the specified sensory attribute then the sample higher in the specified dimension (say A) will be chosen more often as higher in intensity of the specified dimension than the other sample (say B), this is P_{pc}. Mathematically this may be written as Eq. (4.2)

$$H_A : P_{pc} > \frac{1}{2} \tag{4.2}$$

The results of the paired directional (2-AFC) test indicate the direction of the specified difference between the two samples. The sensory specialist must be sure that the two samples *only* differ in the single specified sensory dimension. This is often a problem with sensory discrimination testing of foods because changing one parameter frequently affects many other sensory attributes of the products. For example, removing some of the sugar from a sponge cake will likely make the cake less sweet but it would also affect the texture and the browning of the cake. In this case the directional paired comparison would not be an appropriate discrimination test to use.

4.2.1.2 Difference Paired Comparison (or the Simple Difference Test or the Same/Different Test)

This technique is similar to the triangle and duo–trio tests but it is not often used. It is best used, instead of the triangle or duo–trio test, when the product has a lingering effect or is in short supply and the presentation of three samples simultaneously would not be feasible (Meilgaard et al., 2006). In this case, the experimenter wants to determine whether the two samples differ without specifying the dimension(s) of the potential difference. An example would be if the study involves two sponge cakes, identical in formulation, except for the amount of sugar used. It is likely that the two cakes will differ in sweetness but probably also in texture and crust color.

The panelists are presented simultaneously with the two samples and are asked whether they perceive the samples to be the same or different. See Fig. 4.2 for a sample score sheet. The panelists only need to compare the two samples and decide whether they are similar or different. Humans easily make these types of comparisons and thus the task is relatively easy for the panelists. Thus, the panelists must be trained to understand the task as described by the score sheet but they need not be trained to evaluate specified sensory dimensions. The difference paired comparison method has four possible serving sequences (AA, BB, AB, BA). These sequences should be randomized across panelists with each sequence appearing an equal number of times.

The test is one tailed since the experimenter knows the correct answer to the question asked of each of the panelists, i.e., whether the two samples served to a

Fig. 4.2 Example of a difference paired comparison score sheet.

Date ______________
Name ______________

Please rinse your mouth with water before starting. There are two samples in each of the two paired comparison sets for you to evaluate. Taste each of the coded samples in the set in the sequence presented, from left to right, beginning with Set 1. Take the entire sample in your mouth. NO RETASTING. Are the samples within each set the same or different? Circle the corresponding word. Rinse with water between samples and expectorate all samples and water. Then proceed to the next set and repeat the tasting sequence.

Set				
1	_____	_____	SAME	DIFFERENT
2	_____	_____	SAME	DIFFERENT

specific panelists were the same or different. The alternative hypothesis for the difference paired comparison test states that the samples are perceptibly different and that the population will correctly indicate that the samples are the same or different more frequently than 50% of the time. The mathematical form is

$$H_A : P_{pc} > \frac{1}{2} \tag{4.3}$$

The verbal form of the alternative hypothesis is that the population would be correct (saying that AB and BA pairs are different and that AA and BB pairs are the same) more than half the time. The results of the paired difference test will only indicate whether the panelists could significantly discriminate between the samples. Unlike the paired directional test, no specification or direction of difference is indicated. In other words, the sensory scientist will only know that the samples are perceptibly different but not in which attribute(s) the samples differed. An alternative analysis is presented in the Appendix to this chapter, where each panelist sees an identical pair (AA or BB) and one test pair (AB or BA) in randomized sequence.

4.2.2 Triangle Tests

In the triangle test, three samples are presented simultaneously to the panelists, two samples are from the same formulation and one is from the different formulation. Each panelist has to indicate either which sample is the odd sample or which two samples are most similar. The usual form of the score sheet asks the panelist to indicate the odd sample. However, some sensory specialists will ask the panelist to indicate the pair of similar samples. It probably does not matter which question is asked. However, the sensory specialist should not change the format when re-using panelists since they will get confused. See Fig. 4.3 for a sample score sheet. Similarly to the paired difference test the panelist must be trained to understand the task as described by the score sheet.

The null hypothesis for the triangle test states that the long-run probability (P_t) of making a correct selection when there is no perceptible difference between the samples is one in three (H_0:P_t = 1/3). The alternative hypothesis states that the probability that the underlying population will make the correct decision

Date ______________
Name ______________
Set ______________

Rinse your mouth with water before beginning. Expectorate the water into the container provided. You received three coded samples. Two of these samples are the same and one is different. Please taste the samples in the order presented, from left to right. Circle the number of the sample that is different (odd). Rinse your mouth with water between samples and expectorate all samples and the water.

__________ __________ __________

Fig. 4.3 Example of a triangle score sheet.

when they perceive a difference between the samples will be larger than one in three.

$$H_A : P_t > \frac{1}{3} \tag{4.4}$$

This is a one-sided alternative hypothesis and the test is one tailed. In this case there are six possible serving orders (AAB, ABA, BAA, BBA, BAB, ABB) which should be counterbalanced across all panelists. As with the difference paired comparison, the triangle test allows the sensory specialist to determine if two samples are perceptibly different but the direction of the difference is not indicated by the triangle test. Again, the sensory scientist will only know that the samples are perceptibly different but not in which attribute(s) the samples differed.

4.2.3 Duo–Trio Tests

In the duo–trio tests, the panelists also receive the three samples simultaneously. One sample is marked reference and this sample is the same formulation as one of the two coded samples. The panelists have to pick the coded sample that is most similar to reference. The null hypothesis states that the long-run probability (P_{dt}) of the population making a correct selection when there is no perceptible difference between the samples is one in two (H_0: $P_{dt} = 1/2$). The alternate hypothesis is that if there is a perceptible difference between the samples the population would match the reference and the sample correctly more frequently than one in two times.

$$H_A : P_{dt} > \frac{1}{2} \tag{4.5}$$

Again, the panelists should be trained to perform the task as described by the score sheet correctly. Duo–trio tests allow the sensory specialist to determine if two samples are perceptibly different but the direction of the difference is not indicated by the duo–trio test. In other words, the sensory scientist will only know that the samples are perceptibly different but not in which attribute(s) the samples differed.

There are two formats to the duo–trio test, namely the constant reference duo–trio test and the balanced reference duo–trio test. From the point of view of the panelists the two formats of the duo–trio test are identical (see Figs. 4.4a and b), but to the sensory specialist the two formats differ in the sample(s) used as the reference.

4.2.3.1 Constant Reference Duo–Trio Test

In this case, all panelists receive the same sample formulation as the reference. The constant reference duo–trio test has two possible serving orders (R_A BA, R_A AB) which should be counterbalanced across all panelists. The constant reference duo–trio test seems to be more sensitive especially if the panelists have had prior experience with the product (Mitchell, 1956). For example, if product X is the current formulation (familiar to the panelists) and product Z is a new reformulation then a constant reference duo–trio test with product X as reference would be the method of choice.

4.2.3.2 Balanced Reference Duo–Trio Test

With the balanced reference duo–trio test half of the panelists receive one sample formulation as the

Date ________________
Name ________________

Before starting please rinse your mouth with water and expectorate. There are three samples in each of the two duo–trio sets for you to evaluate. In each set, one of the coded pairs is the same as the reference. For each set taste the reference first. Then taste each of the coded samples in the sequence presented, from left to right. Take the entire sample in your mouth. NO RETASTING. Circle the number of the sample which is most similar to the reference. Do not swallow any of the sample or the water. Expectorate into the container provided. Rinse your mouth with water between sets 1 and 2.

Set			
1	Reference	________	________
2	Reference	________	________

Fig. 4.4a Example of a constant reference duo–trio score sheet.

Fig. 4.4b Example of a balanced reference duo–trio score sheet.

Date ________________
Name ________________

Before starting please rinse your mouth with water and expectorate. There are three samples in each of the two duo–trio sets for you to evaluate. In each set, one of the coded pairs is the same as the reference. For each set taste the reference first. Then taste each of the coded samples in the sequence presented, from left to right. Take the entire sample in your mouth. NO RETASTING. Circle the number of the sample which is most similar to the reference. Do not swallow any of the sample or the water. Expectorate into the container provided. Rinse your mouth with water between sets 1 and 2.

Set			
1	Reference	________	________
2	Reference	________	________

reference and the other half of the panelists receive the other sample formulation as the reference. In this case, there are four possible serving orders (R_A BA, R_A AB, R_B AB, R_B BA) which should be counterbalanced across all panelists. This method is used when both products are prototypes (unfamiliar to the panelists) or when there is not a sufficient quantity of the more familiar product to perform a constant reference duo–trio test.

4.2.4 n-Alternative Forced Choice (n-AFC) Methods

The statistical advantages and hypotheses associated with and the uses of the *n*-AFC tests will be discussed in detail in Chapter 5. As we have seen the 2-AFC method is the familiar directional paired comparison method. The three-alternative forced choice (3-AFC) method is similar to a "directional" triangle method where the panelists receive three samples simultaneously and are asked to indicate the sample(s) that are higher or lower in a specified sensory dimension (Frijters, 1979). In any specific 3-AFC study there are only three possible serving orders (AAB, ABA, BAA or BBA, BAB, ABB) that should be counterbalanced across all panelists. As with the 2-AFC the specified sensory dimension must be the only perceptible dimension in which the two samples may differ. The panelists must be trained to identify the sensory dimension evaluated. They must also be trained to perform the task as described by the score sheet (Fig. 4.5).

The three-alternative forced choice test will allow the sensory scientist to determine if the two samples differ in the specified dimension and which sample is higher in perceived intensity of the specified attribute. The danger is that other sensory changes will occur in a food when one attribute is modified and these may obscure the attribute in question. Another version of the *n*-AFC asks panelists to pick out the weakest or strongest in overall intensity, rather than in a specific attribute. This is a very difficult task for panelists when they are confronted with a complex food system.

4.2.5 A-Not-A tests

There are two types of A-not-A tests referenced in the literature. The first and the more commonly used version has a training phase with the two products followed by monadic evaluation phase (Bi and Ennis, 2001a, b), we will call this the standard A-not-A test. The second version is essentially a sequential paired difference test or simple difference test (Stone and Sidel, 2004), which we will call the alternate A-not-A test. The alternate A-not-A test is not frequently used. In the next section we will discuss the alternate A-not-A test first since the statistical analysis for this version is similar to that of the paired comparison discrimination test. The statistical analyses for the various standard A-not-A tests are based on a different theory and somewhat more complex and will be discussed later.

4.2.5.1 Alternate A-Not-A test

This is a sequential same/difference paired difference test where the panelist receives and evaluates the first

Fig. 4.5 Example of a three-alternative forced choice score sheet.

Date ______________
Name ______________

Please rinse your mouth with water before starting. There are three samples in the set for you to valuate. Taste each of the coded samples in the set in the sequence presented, from left to right. Take the entire sample in your mouth. NO RETASTING. Within the group of three, circle the number of the sweeter sample. Rinse with water between samples and expectorate all samples and water.

________ ________ ________

sample, that sample is then removed. Subsequently, the panelist receives and evaluates the second sample. The panelist is then asked to indicate whether the two samples were perceived to be the same or different. Since the panelists do not have the samples available simultaneously they must mentally compare the two samples and decide whether they are similar or different. Thus, the panelists must be trained to understand the task as described by the score sheet but they need not be trained to evaluate specified sensory dimensions. The alternate A-not-A test, like the difference paired comparison method, has four serving sequences (AA, BB, AB, BA). These sequences should be randomized across panelists with each sequence appearing an equal number of times. The test is one tailed since the experimenter knows the correct answer to the question asked of the panelists namely whether the two samples are the same or different. The null hypothesis of the alternate A-not-A test is the same as the difference paired comparison null hypothesis (H_0: $P_{pc} = 0.5$). The alternative hypothesis for this form of the A-not-A test is that if the samples are perceptibly different the population will correctly indicate that the samples are the same or different more frequently than one in two times. This alternative hypothesis is also the same as that of the difference paired comparison test (H_A: $P_{pc} > 1/2$).

The results of the A-not-A test only indicate whether the panelists could significantly discriminate between the samples when they are not presented simultaneously. Like the paired difference test, no direction of difference is indicated. In other words, the sensory scientist will only know that the samples are perceptibly different but not in which attribute(s) the samples differed.

This version of the A-not-A test is frequently used when the experimenter cannot make the two formulations have exactly the same color or shape or size, yet the color or shape or size of the samples are not relevant to the objective of the study. However, the differences in color or shape or size have to be very subtle and only obvious when the samples are presented simultaneously. If the differences are not subtle the panelists are likely to remember these and they will make their decision based on these extraneous differences.

4.2.5.2 Standard A-Not-A Test

Panelists inspect multiple examples of products that are labeled "A" and usually also products that are labeled "not-A." Thus there is a learning period. Then once the training period has been completed the panelists receive samples one at a time and are asked whether each one is either A or not-A. As discussed by Bi and Ennis (2001a) the standard A-not-A test potentially has four different designs. For the monadic A-not-A test the panelist, after the training phase, is presented with a single sample (either A or not-A). In the paired A-not-A version the panelist, after completion of the training phase, is presented with a pair of samples, sequentially (one A and one not-A, counter balanced across panelists). In the replicated monadic A-not-A version the panelist, after completion of training, receives a series of samples of either A or not-A but not both. This version is rarely used in practice. Lastly, in the replicated mixed A-not-A version the panelist, after completion of training, receives a series of A and not-A samples. Each of these different formats requires different statistical models and using an inappropriate model could lead to a misleading conclusion. As described by Bi and Ennis (2001a) "The statistical models for the A-Not A method are different from that of other discrimination methods such as the *m*-AFC, the triangle, and the duo–trio methods."

"Pearson's and McNemar's chi-square statistics with one degree of freedom can be used for the standard A-Not A method while binomial tests based on the proportion of correct responses can be used for the *m*-AFC, the triangle, and the duo–trio methods. The basic difference between the two types of difference tests is that the former involves a comparison of two proportions (i.e., the proportion of “A” responses for the A sample versus that for the Not A sample) or testing independence of two variables (sample and response) while the latter is a comparison of one proportion with a fixed value (i.e., the proportion of correct responses versus the guessing probability)”. Articles by Bi and Ennis (2001a, b) clearly describe data analysis methods for these tests.. Additionally, the article by Brockhoff and Christensen (2009) describes a R-package called SensR (http://www.cran.r-project.org/package=sensR/) that may be used for the data analyses of some Standard A-not-A tests. The data analyses associated with the standard A-not-A tests are beyond the scope of this textbook, but see the Appendix of this chapter which shows the application of the McNemar chi-square for a simple A-not-A test where each panelist received one standard product (a “true” example of A) and one test product. Each is presented separately and a judgment is collected for both products.

4.2.6 Sorting Methods

In sorting tests the panelists are given a series of samples and they are asked to sort them into two groups. The sorting tests can be extremely fatiguing and are not frequently used for taste and aroma sensory evaluation but they are used when sensory specialists want to determine if two samples are perceptibly different in tactile or visual dimensions. The sorting tests are statistically very efficient since the long-run probability of the null hypotheses of the sorting tests can be very small. For example, the null hypothesis of the two-out-of-five test is 1 in 10 ($P_{2/5} = 0.1$) and for the Harris–Kalmus test the null hypothesis is 1 in 70 ($P_{4/8} = 0.0143$). These tests are discussed below.

4.2.6.1 The Two-Out-of-Five Test

The panelists receive five samples and are asked to sort the samples into two groups, one group should contain the two samples that are different from the other three samples (Amoore et al., 1968). Historically, this test was used for odor threshold work where the samples were very weak and therefore not very fatiguing (Amoore, 1979). The probability of correctly choosing the correct two samples from five by chance alone is equal to 0.1. This low probability of choosing the correct pair by chance is the main advantage of the method. However, major disadvantage of this method is the possibility of sensory fatigue. The panelists would have to make a number of repeat evaluations and this could be extremely fatiguing for samples that have to be smelled and tasted. This technique works well when the samples are compared visually or by tactile methods but it is usually not appropriate for samples that must be smelled or tasted. Recently Whiting et al. (2004) compared the two-out-of-five and the triangle test in determining perceptible differences in the color of liquid foundation cosmetics. They found that the triangle test results gave weak correlations with the instrumental color-differences but that the results of the two-out-of-five test were well correlated with the instrumental values.

4.2.6.2 The Harris–Kalmus Test

The Harris–Kalmus test was used to determine individual thresholds for phenyl thiocarbamide (PTC, a.k.a. phenyl thiourea, PTU). In this test panelists are exposed to increasing concentration levels of PTC in groups of eight (four samples containing water and four samples containing the current concentration of PTC). The panelists are asked to sort the eight samples into two groups of four. If the panelist does the sorting task incorrectly he/she is then exposed to the next higher concentration of PTC. The sorting task continues until the panelist correctly sorts the two groups of four samples. That concentration level of PTC is then identified as the threshold level for that panelist (Harris and Kalmus, 1949–1950). The method has the same disadvantage as the two-out-of-five test, in that it could be fatiguing. However, as soon as the panelist correctly sorts the samples the researcher concludes that the panelist is sensitive to PTC. Panelists insensitive to PTC only “taste” water in the solutions and are thus not fatigued. A shortened version of this test using three-out-of-six was used by Lawless (1980) for PTC and PROP (6-*n*-propyl thiouracil) thresholds.

4.2.7 The ABX Discrimination Task

The ABX discrimination task, as its name intends to suggest, is a matching-to-sample task. The panelist receives two samples, representing a control sample and a treatment sample. As in other discrimination tasks, the "treatment" in food research is generally an ingredient change, a processing change or a variable having to do with packaging or shelf life. The "X" sample represents a match to one of the two inspected samples and the panelist is asked to indicate which one is the correct match. The chance probability level is 50% and the test is one tailed, as the alternative hypothesis is performance in the population above 50% (but not below). In essence, this task is a duo–trio test in reverse (Huang and Lawless, 1998). Instead of having only one reference, two are given, as in the dual standard discrimination test. In theory, this allows the panelists to inspect the two samples and to discover for themselves the nature of the sensory difference between the samples, if any. As the differences are completely "demonstrated" to the panelists, the task should enjoy the same advantage as the dual standard test (O'Mahony et al., 1986) in that the participants should be able to focus on one or more attributes of difference and use these cues to match the test item to the correct sample. The inspection process of the two labeled samples may also function as a warm-up period. The test may also have some advantage over the dual standard test since only one item, rather than two are presented, thus inducing less sensory fatigue, adaptation, or carry-over effects. On the other hand, giving only one test sample provides less evidence as to the correct match, so it is unknown whether this test would be superior to the dual standard. As in other general tests of overall difference (triangle, duo–trio) the nature of the difference is not specified and this presents a challenge to the panelists to discover relevant dimensions of sensory difference and not be swayed by apparent but random differences. As foods are multi-dimensional, random variation in irrelevant dimensions can act as a false signal to the panelists and draw their attention to sensory features that are not consistent sources of difference (Ennis and Mullen, 1986).

This test has been widely used as a forced choice measure of discrimination in psychological studies, for example, in discrimination of speech sounds and in measuring auditory thresholds (Macmillan et al., 1977; Pierce and Gilbert, 1958). Several signal detection models (see Chapter 5) are available to predict performance using this test (Macmillan and Creelman, 1991). The method has been rarely if ever applied to food testing, although some sensory scientists have been aware of it (Frijters et al., 1980). Huang and Lawless (1998) did not see any advantages to the use of this test over more standard discrimination tests.

4.2.8 Dual-Standard Test

The dual standard was first used by Peryam and Swartz (1950) with odor samples. It is essentially a duo–trio test with two reference standards—the control and the variant. The two standards allow the panelists to create a more stable criterion as to the potential difference between the samples. The potential serving orders for this test are $R_{(A)}$ $R_{(B)}$, AB, $R_{(A)}$ $R_{(B)}$ BA, $R_{(B)}$ $R_{(A)}$ AB, $R_{(B)}$ $R_{(A)}$ BA. The probability of guessing the correct answer by chance is 0.5 and the data analyses for this test are identical to that of the duo–trio test. Peryam and Swartz felt quite strongly that the technique would work best with odor samples due to the relatively quick recovery and that the longer recovery associated with taste samples would preclude the use of the test. The test was used by Pangborn and Dunkley (1966) to detect additions of lactose, algin gum, milk salts, and proteins to milk. O'Mahony et al. (1986) working with lemonade found that the dual-standard test elicited superior performance over the duo–trio test. But O'Mahony (personal communication, 2009) feels that this result is in error, since the panelists were not instructed to evaluate the standards prior to each pair evaluation and therefore the panelists were probably reverting to a 2-AFC methodology. This would be in agreement with Huang and Lawless (1998) who studied sucrose additions to orange juice and they did not find superiority in performance between the dual standard and the duo–trio or the ABX tests.

4.3 Reputed Strengths and Weaknesses of Discrimination Tests

If the batch-to-batch variation within a sample formulation is as large as the variation between formulations