

C H A P T E R

6

Signal Detection Measures

6.1 Introduction

- 6.1.1 What is Signal Detection Theory (SDT)?
- 6.1.2 A Recap on Some Terminology: N , m and M
- 6.1.3 Why Measure d' ?

6.2 Section A: Practice

- 6.2.1 Signal Detection Theory with Palamedes
- 6.2.2 Converting P_c to d' for Unbiased M-AFC Tasks
- 6.2.3 Measuring d' for 1AFC Tasks
- 6.2.4 Measuring d' for 2AFC Tasks with Observer Bias
- 6.2.5 Measuring d' for Same-Different Tasks
- 6.2.6 Measuring d' for Match-to-Sample Tasks
- 6.2.7 Measuring d' for M-AFC Oddity Tasks
- 6.2.8 Estimating $P_{c_{\max}}$ with Observer Bias

6.2.9 Comparing d 's and P_c s across Different Tasks

6.3 Section B: Theory

- 6.3.1 Relationship Between Z-scores and Probabilities
- 6.3.2 Calculation of d' for M-AFC
- 6.3.3 Calculation of d' and Measures of Bias for 1AFC Tasks
- 6.3.4 Calculation of d' for Unbiased and Biased 2AFC Tasks
- 6.3.5 Calculation of d' for Same-Different Tasks
- 6.3.6 Calculation of d' for Match-to-Sample Tasks
- 6.3.7 Calculation of d' for M-AFC Oddity Tasks

Further Reading

Exercises

References

6.1 INTRODUCTION

6.1.1 What is Signal Detection Theory (SDT)?

In performance-based psychophysical tasks there are many situations in which proportion correct (P_c) is inappropriate, uninformative, or invalid. In this chapter we will examine some of these situations and describe a popular alternative measure termed d' ("d-prime"). d' is a measure derived from a branch of psychophysics known as signal detection theory (SDT). In Chapter 4 Section 4.3.1.2, SDT was introduced as one of the models of how perceptual decisions were made in a forced-choice task. The SDT model attempted to explain the shape of the psychometric function relating P_c to stimulus magnitude. It was argued that the presence of internal noise, or uncertainty, led to stimuli being represented in the brain not by a single point along a sensory continuum, but as a random sample drawn from a distribution with a mean and a variance. SDT is therefore a theory of how observers make perceptual decisions, given that the stimuli are represented stochastically (or probabilistically) inside the brain. The purpose of this chapter is to discuss why d' is a useful measure of performance, to describe the Palamedes routines that convert conventional measures of performance such as P_c into d' , and to explain the theory behind those conversions.

SDT is a large topic and it is impossible to do it justice in a single chapter of an introductory book on psychophysics. There are a number of excellent books and articles on SDT (see Further Reading at the end of the chapter) that cover a wider range of material and provide a more in-depth treatment than is possible here. In particular, Macmillan and Creelman's (2005) comprehensive treatment of SDT is strongly recommended as an accompaniment to this chapter. Although our chapter is modest by comparison to textbooks specializing in SDT, it nevertheless aims to do more than just scratch the surface of the topic. The Palamedes SDT routines described in Section A of the chapter are generous in terms of the number of psychophysical procedures they cover. Section A is intended to familiarize readers with the basic concepts involved and the practical tools necessary for converting psychophysical measurements into d' s, and *vice versa*, without the need to understand the underlying theory. The theory is provided in Section B, and we have attempted to make it as accessible as possible.

6.1.2 A Recap on Some Terminology: N, m and M

As elsewhere in this book, we use the term "forced-choice" for any task in which observers are required on each trial to make a forced-choice response, irrespective of the number of stimuli or stimulus alternatives presented on the trial. We denote the number of stimuli presented during a trial as N . m refers to the number of response choices available to the observer, and as we saw in Chapter 4 this determines the

expected chance performance or guessing rate of the task, calculated as $1/m$. Remember that m and N are not always the same. For example, in a yes/no task N is 1 because only one of two stimulus states is presented per trial, target-present or target-absent, but m is 2 because there are two choices of response – "yes" or "no." In a same-different task N can be 2 or 4, depending on whether the Same and Different pairs are presented on separate trials or together in the same trial. When the pairs are presented on separate trials the task for the observer is to respond "same" or "different," whereas when presented together in the same trial the task is to respond "1" or "2," depending on which interval contains the Different (or Same) pair. However, for both varieties of same-different, m is 2.

For the purposes of SDT the third parameter described in Chapter 2, M , is especially important. M is the "number of stimulus alternatives presented per trial." In the $N = 4$ same-different task described above, M is 2, as two stimulus alternatives are presented per trial – the Same pair and the Different pair. Although m is also 2 for this task, m and M are not always the same. For example, with the yes/no task m is 2 but M is 1, since although there are two available response choices, only one stimulus alternative is presented on a trial, either "target present" or "target absent." As we stated in Chapter 2, forced-choice tasks in this book are prefixed with the value of M . Table 6.1 summarizes the relationship between N , m , and M for the main varieties of task discussed in this chapter.

TABLE 6.1 Relationship between M , N , and m for the psychophysical tasks discussed in the chapter

Task	Acronym prefixed by number of stimulus alternatives per trial M	Number of stimuli per trial N	Number of response options per trial m
Yes/no	1AFC	1	2
Symmetric single-alternative	1AFC	1	2
Standard two alternative forced-choice	2AFC	2	2
Single alternative same-different	1AFC	2	2
Two alternative same-different	2AFC	4	2
Two alternative match-to-sample	2AFC	3	2
Three alternative oddity	3AFC	3	3
Standard M -alternative forced-choice	M -AFC	M	M
M -alternative match-to-sample	M -AFC	$M + 1$	M
M -alternative oddity	M -AFC	M	M

One could argue that because this chapter deals exclusively with forced-choice tasks, prefixing every task with the acronym AFC is unnecessary because it is redundant, and that instead AFC should be used sparingly as in standard SDT texts (e.g., Macmillan & Creelman, 2005). However, in a book dealing with psychophysical procedures that are not all forced-choice, the acronym helps make explicit those that are.

Recall, also, that for most tasks the stimulus alternatives can be presented in spatial or temporal order, and hence be denoted as AFC or IFC. AFC, however, is the generic acronym, so we have adopted this as the default and used IFC only when specifically referring to tasks in which the stimulus alternatives are presented in temporal order. From the point of view of signal detection theory, however, the two acronyms are interchangeable.

6.1.3 Why Measure d' ?

Suppose we wanted to compare the results of two texture-segregation experiments, one that employed a standard 4AFC task and the other a standard 2AFC task. The stimuli employed in texture segregation experiments typically consist of a "target" texture embedded in a "background" texture. The target and background differ in some textural property, such as the average orientation or average size of the texture elements. In the popular 4AFC version, the target is embedded in one of four quadrants of the background texture, and on each trial the observer selects the quadrant containing the target. In the 2AFC version, the embedded target is typically positioned on one or other side of the fixation point. In the 4AFC task proportion correct (P_c) would be expected to range from 0.25–1, since the guessing rate $1/m$ is 0.25. For the 2AFC version P_c would be expected to range from 0.5–1, since the guessing rate is 0.5. Yet, presumably, the underlying sensory mechanisms involved in segregating the target from the background are the same for both the 2AFC and 4AFC tasks, especially if the target locations are arranged equidistant from fixation in order to stimulate equivalent visual mechanisms. Thus, any differences in performance between the two tasks, which will be most prominent when performance is close-to-chance, are unlikely to be due to differences in the observer's sensitivity to the stimuli, but more probably due to differences in the uncertainty of the target location. Put another way, with four possible target locations the observer is more likely to make a mistake than with two target locations, all else being equal. And the more possible target locations, the more mistakes the observer will likely make. One reason for using d' is that it can remove, or take into account, the effects of target location uncertainty, providing a measure of performance that is procedure-free. In other words, for M -AFC tasks, d' may equate performance across M . We stress "may" because it is ultimately an empirical, not a theoretical, question as to whether d' does equate performance across M , and there are some situations where it has been shown not to do so (e.g., Yeshurun, Carrasco, & Maloney, 2008).

Although the most popular value of M in forced-choice tasks is 2 (either 2AFC or 2IFC), M can be much higher. For example, one of the authors once conducted experiments in which observers were required to choose a column of pixels with the highest average intensity from 256 columns (see Kingdom, Moulden, & Hall 1987) for details). So M for this task was 256!

Another reason for using d' is that in some situations it removes the inherently sigmoidal- or bow-shape of the psychometric function when plotted in terms of P_c . Again, however, it is an empirical question as to whether d' does linearize the psychometric function, and in theory it will only do so if the relationship between sensory magnitude and stimulus magnitude is linear, and the internal noise levels remain constant with stimulus magnitude. However, if these constraints are satisfied Figure 6.1 illustrates how d 's equate performance across M and linearize the psychometric function. The figure shows hypothetical P_c s as a function of stimulus level for a standard 2AFC, 4AFC, and 8AFC task, where m is respectively 2, 4, and 8 (see above), under the assumption that observer sensitivity for each stimulus magnitude is the same across tasks. Note that as m increases the functions extend downwards, since chance performance, $1/m$, decreases. When the data are replotted as d 's, however, the functions straighten and superimpose.

The linearizing effect of d' is particularly useful when one wants to correlate psychophysical performance with some other dimension. For example, suppose one wanted to study the effects of age on the ability to discriminate the speed of moving vehicles in seniors, using movies of moving vehicles as test stimuli. If insufficient time was available to test each observer with enough stimuli to derive full psychometric functions relating P_c to speed difference, a single P_c at a particular speed

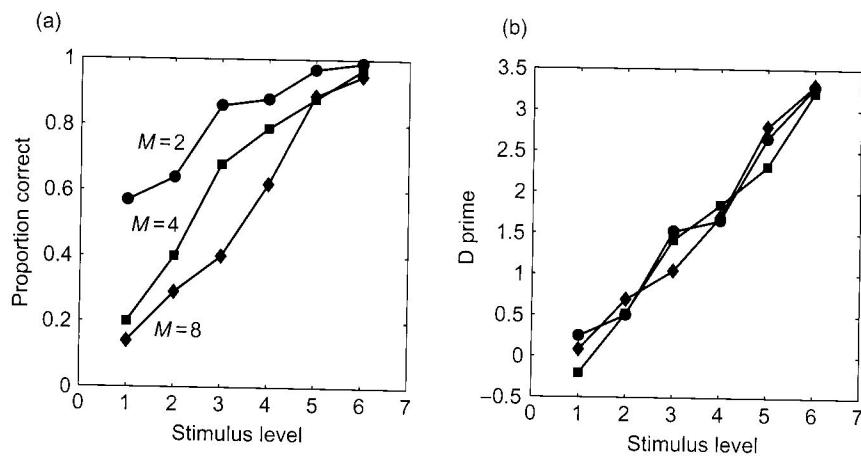


FIGURE 6.1 (a) Hypothetical P_c (proportion correct) data for an $M = 2$, $M = 4$, and $M = 8$ forced-choice task. (b) The same data plotted as d 's.

difference might have to suffice. Given the likely range of performance across observers of the same age and across ages, converting the P_{Cs} into d 's may well linearize the data and render the measured (likely negative) correlation between performance and age more valid. Moreover, correlation is not the only type of statistical manipulation potentially benefitting from converting P_{Cs} into d 's. Parametric tests such as t-tests and Analysis-of-Variance (ANOVA), which are often employed to make inferences about group differences, are only valid if the data within each group are normally distributed. Converting P_{Cs} to d 's may turn out to satisfy this requirement.

A third, and for many investigators the most important, reason for using d 's is that certain types of psychophysical task are prone to the effects of observer bias. In Chapters 2 and 3 we noted for example, that in the yes/no task observers tend to be biased towards responding "yes" or "no," irrespective of their underlying sensitivity to the target. As we shall see, the greater the bias, the smaller the expected P_c , making P_c an invalid measure of sensitivity. d' takes into account the effects of bias, thus providing a relatively "bias-free" measure of performance.

6.2 SECTION A: PRACTICE

6.2.1 Signal Detection Theory with Palamedes

Palamedes contains a large number of routines for performing signal detection computations. To understand the convention for the names of the routines, consider the following example: **PAL_SDT_2AFCmatchSample_DiffMod_PCToDP**. The routine is identifiable as part of the SDT package by the prefix **PAL_SDT**. The term **2AFCmatchSample** identifies the task, i.e., match-to-sample, and the value of M , i.e., 2. The generic acronym AFC is used in all routines. The term **DiffMod** specifies the particular model for the computation; in this case it means a differencing model. By "model" we mean the particular strategy that the observer is assumed to adopt when performing the task. This is only relevant to those tasks for which there is more than one possible strategy. Finally **PCToDP** specifies the actual computation, in this case the conversion of P_c to d' . This last term is invariably of the form XtoY, where X is the principle input argument(s) and Y the principle output argument(s). However, the routine may require additional input arguments and may output additional arguments. The abbreviations used for the XtoY term are **PC** for P_c , **DP** for d' , **PHF** for proportion hits and proportion false alarms, and **PH** for proportion hits. If the required input to a routine is either **PC** or **DP**, the data can be entered as a scalar, vector or matrix. If the input is **PHF** the data must be an $m \times 2$ (rows \times columns) matrix with a minimum of one row; one column is for the proportion of hits and the other column for the corresponding proportion of false alarms.

There are two important assumptions underlying all the routines described in this chapter. The first concerns the stimuli. We assume that all the stimulus alternatives

are presented the same number of times in each experiment. Thus, in a yes/no task we assume that there are as many target-present as target-absent trials, and in a same-different task as many Same pairs as Different pairs. The second assumption is that the observer's internal noise variance is the same across stimulus alternatives. This is the "default" assumption for most SDT tasks. There are procedures for determining from data whether this assumption is violated, but they will not be dealt with here and the interested reader is referred to Macmillan and Creelman (2005) for the necessary details.

6.2.2 Converting P_c to d' for Unbiased M-AFC Tasks

We begin with the routine that deals with the standard M-AFC task, where M is any value greater than 1. For this class of task the three variables, N , M , and m , are equal. Although tables exist for converting P_c to d' for a range of M (Elliot, 1964; Macmillan & Creelman, 2005) the Palamedes routines work for any value of M and are simple to implement.

In an M -AFC task, one of the alternatives on each trial contains the target, while the remaining $M-1$ alternatives contain no target. If the observer selects the alternative containing the target their response is scored "correct," otherwise it is "incorrect," and P_c is calculated as the proportion of trials in which the observer is scored correct.

The two routines for standard M-AFC are **PAL_SDT_MAFc_DPttoPC**, which converts d' to P_c , and **PAL_SDT_MAFc_PctoDP**, which converts P_c to d' . The routines make an important assumption in addition to those described in the previous section. The assumption is that the observer is not biased to respond to any one alternative/interval more than any other. If the assumption is not true and the observer is biased, then the estimates of d' will not be close to the "true" values.

Each routine takes two arguments. The first is a scalar, vector or matrix of the measure to be converted (d' or P_c), and the second is the value of M where $M > 1$. Typically one wants to convert P_c s into d' s, so try filling a vector named **PropCorr** with an array of P_c s as follows:

```
>> PropCorr = [.3:.1:.9];
```

To convert the array to d' 's for, say, a 3AFC task, type and execute:

```
>> DP = PAL_SDT_MAFc_PctoDP(PropCorr, 3)
```

The array returned is:

```
DP =
-0.1207 0.2288 0.8852 1.6524 2.2302
```

Note that the first value in the array is negative. d' is zero when performance is at chance, which for the standard 3AFC task is 0.33, so any P_c below 0.33 will

produce a negative d' . Try repeating the above with $M = 4$. Note that the first value is now positive, since chance level for a 4AFC task is 0.25. If one sets M to 2 (chance = 0.5) the first two d 's are negative. One can see from these examples that increasing M for a given P_c increases d' . This is because as M increases so too does the chance that one of the non-target intervals/locations will contain a signal that is by chance greater in magnitude than the interval/location containing the target. In other words, for a given P_c , observer sensitivity is computed to be higher if the task has a large compared to a small M .

Try also converting an array of d 's to P_c s using **PAL_SDT_MAFC_DPtOPC**. Note that increasing M for a given d' this time decreases P_c , because the more possible target intervals/ locations, the more likely one of them will, by chance, contain a signal greater than the interval/location containing the target.

6.2.3 Measuring d' for 1AFC Tasks

6.2.3.1 d' from pH and pF

As discussed in Chapters 2 and 3, 1AFC tasks, especially those that are not symmetric such as the yes/no task, are particularly prone to bias. Remember that with the yes/no task, the observer is required to indicate on each trial whether the target stimulus is present or absent. If the observer adopts a loose criterion, this will result in a bias towards responding "yes," whereas adopting a strict criterion will result in a bias towards responding "no." Both types of bias may occur irrespective of how sensitive the observer is to the stimulus. For this reason, signal detection theory approaches the computation of d' for 1AFC tasks differently from tasks that are assumed to be bias-free. Rather than use P_c , the responses from a 1AFC task are divided into two groups: the target-present trials in which the observer responds "yes" (i.e., correctly) and the target-absent trials in which the subject responds "yes" (i.e., incorrectly). The former responses are commonly termed "hits," the latter "false alarms." The proportion of target-present trials that are hits is given here by pH , and the proportion of target-absent trials that are false alarms, pF . Note that the overall P_c is given by $[pH + (1 - pF)]/2$, since $1 - pF$ gives the proportion of target-absent trials in which the observer responds "no," i.e., also correctly. The two measures pH and pF can be used not only to calculate d' , but also to calculate the bias towards responding "yes" or "no." The calculations are explained in Section B.

The Palamedes routine that converts pH and pF to d' , as well as to two measures of bias, is **PAL_SDT_1AFC_PHFToDP**. The input argument can either be a pre-named $m \times 2$ matrix of pH and pF values, or the raw values themselves. There are four output arguments: d' ; two measures of bias termed C and $\ln\beta$; and overall P_c .

Suppose we want to input just a single pair of raw pH and pF values. Type and execute the following, and remember to place the square brackets around the two values so that they are entered as a matrix:

```
>> [dp C lnB Pc] = PAL_SDT_1AFC_PHFToDP([0.6 0.1])
```

The output should be:

```
dp =
1.5349
C =
0.5141
lnB =
0.7891
Pc =
0.7500
```

The criterion C can range from negative to positive, with negative values indicating a bias towards "yes" and positive values a bias towards "no." The criterion measure $\ln\beta$ shows the same pattern as C . The positive values of $C = 0.51$ and $\ln\beta = 0.78$ in the above example are indicative of a relatively strict criterion, that is a bias towards responding "no."

To explore the relationship between d' , bias, pH , and pF , one can also use the reverse routine **PAL_SDT_1AFC_DPtоСHF**. For example, create a vector named **dprime** filled with a $\times 5$ array of 2s, and a vector named **criterion** with values $-1, -0.5, 0, 0.5$, and 1 . Then type and execute:

```
>> pHF = PAL_SDT_1AFC_DPtоСHF(dprime,criterion)
```

The output should be:

```
pHF =
0.9772 0.5000
0.9332 0.3085
0.8413 0.1587
0.6915 0.0668
0.5000 0.0228
```

The first column gives pH , the second pF . Note that as C increases (loose to strict criterion) both the number of hits and the number of false alarms decreases. Figure 6.2a shows the relationship between pH and pF as a function of C for three values of d' . As one travels along each of the curves from left to right, C is decreasing, resulting in an increase in both pH and pF . The relationship between pH and pF is known as a receiver operating characteristic, or ROC. The ROC in Figure 6.2a is hypothetical, but ROCs can be generated from experiments in which the responses are not binary options such as "yes" or "no," but ratings, for example 1 to 5, as to how confident one is that the target is present. Currently Palamedes does not provide routines for analyzing rating-scale data, but the method, along with the value of ROCs for testing the equal-variance assumption mentioned above, is described in Macmillan and Creelman (2005).

Figure 6.2.b shows why Pc is not a good measure of performance when there is bias. Assuming that the "true" observer's sensitivity is given by d' , one can see that

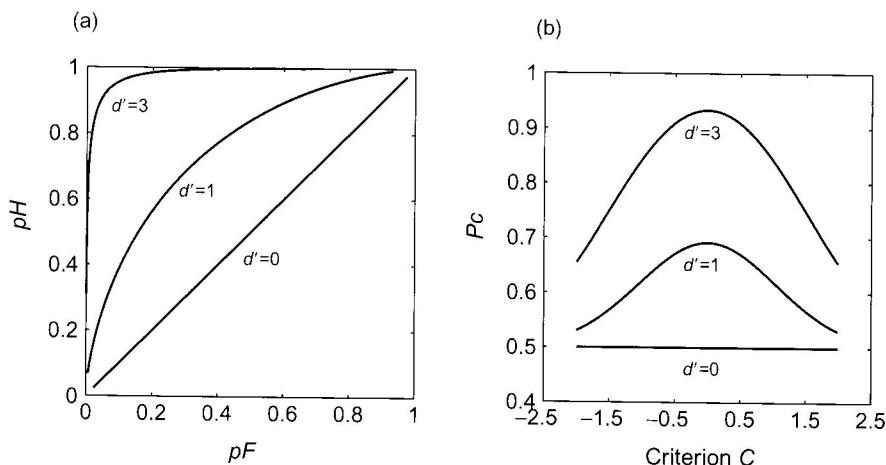


FIGURE 6.2 (a) Hypothetical receiver operating characteristics, or ROCs, for three d' 's. Note that pH (proportion of hits) is plotted against pF (proportion of false alarms). As the criterion C decreases, one moves along each curve from left to right. (b) the effect of changing C on the overall P_c (proportion correct) for the same d' 's as in (a). Note that all three curves peak when $C = 0$.

P_c varies considerably with criterion C . Only if there is no bias ($C = 0$) is P_c a valid measure of performance. A zero-bias assumption may sometimes be reasonable with "symmetric" 1AFC tasks, such as the 1AFC orientation discrimination experiment discussed in Chapter 3. However, some researchers argue that even with symmetric 1AFC tasks the data should be analyzed under the presumption that bias might occur. If it turns out there is no bias then nothing is lost, but if bias is found to occur it is taken into account – a win-win situation.

How, then, do we convert the responses from the orientation discrimination experiment into pH and pF ? The answer is to classify the responses in a way analogous to that of the yes/no experiment. For the orientation discrimination experiment this means classifying a "left-oblique" response as a "hit" when the stimulus is left-oblique, and as a "false alarm" when the stimulus is right-oblique. pH is then the proportion of "left-oblique" responses for the left-oblique stimuli, and pF the proportion of "left-oblique" responses for the right-oblique stimuli. Note that pH and pF defined in this way are sufficient to describe all the responses in the experiment i.e., including the "right-oblique" responses. The proportion of times the observer responds "right-oblique" is $1 - pH$ for the left-oblique stimulus trials and $1 - pF$ for the right-oblique stimulus trials. Note also that as with yes/no, overall P_c is given by $[pH + (1 - pF)]/2$. Thus, if the observer in the orientation discrimination experiment is biased towards responding "left-oblique," both pH and pF , as defined above, will tend to be relatively high, and by comparing the two in the same way as with the yes/no task the bias can be taken into account and a valid measure of sensitivity calculated.

6.2.3.2 1AFC Demonstration Programs

The program **PAL_SDT_1AFC_PHFtoDP_Demo** illustrates how the routines for 1AFC tasks can be incorporated into a program that generates a more user-friendly output of d' and criterion measures. When executed, the program prompts you as follows:

Enter a matrix of proportion Hits and False Alarms

You must enter arrays of raw values. An example input matrix of pH and pF values would be:

[0.6 0.2; 0.7 0.2; 0.8 0.2]

The output should be:

pH	pF	Dprime	propCorr	Crit C	lnBeta
0.6000	0.2000	1.0950	0.7000	0.2941	0.3221
0.7000	0.2000	1.3660	0.7500	0.1586	0.2167
0.8000	0.2000	1.6832	0.8000	-0.0000	-0.0000

The inverse routine **PAL_SDT_1AFC_DPtоНHF_Demo** operates similarly. You are prompted for two vectors of numbers. Try entering the following values, then execute:

Enter a vector of Dprime values [1 2 3]

Enter a vector of Criterion C values [0.5 0.5 0.5]

The output should be:

dprime	critC	pH	pF	pCorr
1.0000	0.5000	0.5000	0.1587	0.6707
2.0000	0.5000	0.6915	0.0668	0.8123
3.0000	0.5000	0.8413	0.0228	0.9093

6.2.4 Measuring d' for 2AFC Tasks with Observer Bias

Although the inherent symmetry of 2AFC tasks makes them less susceptible to bias than the yes/no task, a bias towards responding to one alternative/interval more than the other may still occur, and if it does occur P_c becomes an invalid measure of sensitivity. As with symmetric 1AFC tasks, some researchers prefer not to hedge their bets with 2AFC and analyze the data on the presumption that bias might have occurred.

To take into account bias in 2AFC tasks, the observer's responses need to be classified as hits and false alarms, as with the symmetric 1AFC task. Let the response be "1" or "2" depending on the alternative perceived to contain the target. A "1" response is designated as a "hit" when the target is present in the first alternative/interval, and as a "false alarm" when the target is present in the second alternative/interval. Thus pH is the proportion of "1" responses for targets presented in the first alternative/interval and pF the proportion of "1" responses for targets presented in the second alternative/interval. Note that, as with 1AFC tasks, pH and pF defined in this way are sufficient to describe the full pattern of responses. Thus $1-pH$ is the proportion of "2" responses for targets presented in the first alternative/interval and $1 - pF$ the proportion of "2" responses for targets presented in the second alternative/interval. Note also that, as with 1AFC tasks, overall P_c is given by $[pH + (1 - pF)]/2$.

Palamedes has two routines for the standard 2AFC task when the input arguments are pH and pF : **PAL_SDT_2AFC_DPtOPH** and **PAL_SDT_2AFC_PHFtoDP**. The input and output arguments correspond to those for the 1AFC routines. Remember that one can also use **PAL_SDT_MAFC_PCToDP** and **PAL_SDT_MAFC_DPtOPC** for 2AFC tasks (by inputting data in the form of P_c and setting the argument M to 2), but only if one is happy to assume that the observer is unbiased.

What is the expected relationship between performance in a 1AFC and 2AFC task? One can use **PAL_SDT_1AFC_PHFtoDP** and **PAL_SDT_2AFC_PHFtoDP** to find out. Try the following. Input the same pair of pH and pF values and the same value of the criterion to both routines. Take the ratio of the resulting 1AFC to 2AFC d' 's. The result should be $\sqrt{2}$. The $\sqrt{2}$ relationship between d' 's for 1AFC and 2AFC is often emphasized in expositions of SDT, but one must be careful with its interpretation. It is tempting to suppose that if one performed a 1AFC task and a 2AFC task using the same stimulus magnitudes, the computed d' 's would likely come out in a ratio of $\sqrt{2}$. In fact, the d' 's would likely be very similar. Remember that d' is a measure of sensitivity that is ostensibly independent of the method used to obtain it (although be reminded of the cautionary note from Yeshurun et al., 2008). The likely difference between the two tasks will be in P_c , not d' . As Figure 6.3 demonstrates, the same d' predicts different P_c s for 1AFC and 2AFC. Put another way, observers will typically find a 1AFC task more difficult than a 2AFC task for the same stimulus magnitudes. This is because there is more information in a 2AFC compared to 1AFC trial.

6.2.5 Measuring d' for Same-Different Tasks

In Chapters 2 and 3 we described the class of psychophysical task termed "same-different." One reason for using same-different tasks is that the observer is not required to know the basis on which the discriminands differ. There are two main varieties of same-different task. In the 1AFC version only one pair, Same or Different, is presented on a trial, and the observer has to decide "same" or "different." The pair can be presented either together on the display or in temporal order.

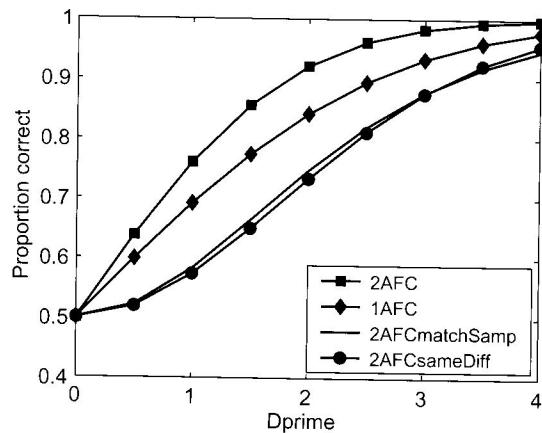


FIGURE 6.3 Example output from **PAL_SDT_DPttoPCcomparison_Demo**.

When using **PAL_SDT_PctoDPcomparison_Demo** do not enter any 1s as this will result in a d' of infinity which cannot be plotted. Instead an example vector input could be **[.5:.025:0.98]**.

In the 2AFC version the Same and Different pairs are both presented on a trial (either together on the display or in temporal order), and the observer chooses the alternative/interval containing the Different (or the Same) pair. The 2AFC same-different task is probably the more popular of the two versions in vision experiments, because it is less prone to bias.

6.2.5.1 d' for 2AFC Same-Different

The Palamedes routines for the 2AFC same-different task are **PAL_SDT_2AFCsameDiff_DPttoPC** and **PAL_SDT_2AFCsameDiff_PctoDP**. Both routines assume an unbiased observer that adopts the strategy of selecting the pair with the greater (or smaller) absolute perceived difference. The routines implement the equations in Macmillan, Kaplan, and Creelman (1977) for a "4IAX" same-different task, where 4IAX denotes that the four stimuli are presented in temporal order, the more typical scenario in an auditory experiment. Both of the Palamedes routines take a single argument (d' or P_c) and output a single argument (P_c or d'). The input arguments may be scalars, vectors or matrices.

6.2.5.2 d' for 1AFC Same-Different

For same-different tasks where only one pair, Same or Different, is presented in a trial, Macmillan and Creelman (2005) argue that observers typically adopt one of two strategies: the "independent observation" or "differencing" strategy (note that Macmillan and Creelman occasionally refer to the 1AFC same-different task as 2IAX or AX. The first acronym denotes that the two stimuli are presented in different temporal intervals).

Suppose that during a session there are only two stimuli: S_1 and S_2 . On each trial the observer is presented with one of four possible combinations: $\langle S_1S_1 \rangle$, $\langle S_2S_2 \rangle$, $\langle S_1S_2 \rangle$, or $\langle S_2S_1 \rangle$. Macmillan and Creelman argue that the most likely strategy in this scenario is that the observer independently assesses the likelihood that each stimulus in a pair is either S_1 or S_2 . The decision "different" is made when the joint likelihood of the pair being S_1 and S_2 exceeds the observer's criterion. This is the independent observation strategy.

The differencing strategy is less optimal, but under some circumstances the more likely to be adopted. As with the strategy assumed for the 2AFC same-different task described above, the decision rule is based on the perceived *difference* between the two stimuli in each pair. The observer responds "different" when the absolute perceived difference between the two stimuli exceeds the criterion. According to Macmillan and Creelman, the differencing strategy is more likely to be adopted when many different stimuli are presented during a session, termed a "roving" experiment. For example, suppose that one wished to compare the detectability of four types of color manipulation applied to images of natural scenes. Let the four manipulations be shifts in average color towards either red, green, blue or yellow. On each trial observers are presented either with two identical natural-scene images (the Same pair) or two images in which the average color of one of the pair was shifted towards one of the four (randomly selected) colors (the Different pair). It would be difficult for observers to independently assess the likelihood that each member of a pair had been subject to a particular color shift, because there are four possible types of color shift. The more likely strategy in this situation would be that observers assess the difference in color between the images in each pair and base their decision accordingly.

6.2.5.2.1 d' for 1AFC Same-Different: Independent Observation Model

The Palamedes routines for the 1AFC same-different task that assumes an independent-observation model are **PAL_SDT_1AFCsameDiff_IndMod_PHFtoDP** and **PAL_SDT_1AFCsameDiff_IndMod_DPtoPHF**. The first routine takes two arguments, a $m \times 2$ matrix of pHs and pFs , and outputs two arguments: d' and criterion C. The second routine performs the reverse operation. For example, try inputting the same matrix of pHs and pFs as for the basic 1AFC task described earlier, i.e.:

```
PHF = [0.6 0.2; 0.7 0.2; 0.8 0.2]
```

then type and execute:

```
>> [dp C] = PAL_SDT_1AFCsameDiff_IndMod_PHFtoDP(PHF)
```

The output should be three d' and three criterion C values. Compare these with those obtained using **PAL_SDT_1AFC_PHFtoDP**. Try also computing d' 's for a bias-free version of the same-different task by setting pF equal to $1 - pH$. You will see that the resulting d' 's under the independent observation model are the same as

those for the 2AFC same-different task ($P_c = pH$), which assumes a differencing strategy.

6.2.5.2.2 d' for 1AFC Same-Different Tasks: Differencing Model

The Palamedes routines for the 1AFC same-different task assuming the differencing model are **PAL_SDT_1AFCsameDiff_DiffMod_PHFtoDP** and **PAL_SDT_1AFCsameDiff_DiffMod_DPttoPHF**. They are implemented in the same way as the routines for the independent observer model. However, they return a different measure of bias termed k (Macmillan & Creelman, 2005). Unlike C , k is not zero when the observer is unbiased.

Consider the following. For a given pH and pF , would you expect d' to be larger or smaller for the differencing compared to the independent observer model? Try various pH and pF combinations to test your predictions.

6.2.6 Measuring d' for Match-to-Sample Tasks

In a match-to-sample task, the observer is presented with a "Sample" stimulus followed by two or more "Match" stimuli, one of which is the same as the Sample – the one the observer must choose. Match-to-sample procedures are particularly popular in animal research, research into children's perception, and studies of cognitive vision (see Chapter 3). As with the same-different task, one advantage of match-to-sample over standard M-AFC is that the observer need not know the basis on which the discriminands differ. The minimum number of stimuli per repeat trial in a match-to-sample task is three (one Sample; two Match), and this is undoubtedly the most popular design. With two Match stimuli the task is 2AFC according to our naming system. Macmillan and Creelman (2005) refer to the task as ABX.

6.2.6.1 d' for 2AFC Match-to-Sample

Macmillan and Creelman argue that for the ABX task, observers may adopt either independent observation or differencing strategies, the latter more likely in "roving" experiments where a number of different stimulus pairs are presented during a session. The independent observation strategy is analogous to that for the same-different task. When adopting the differencing strategy the observer selects the Match that is perceived to be *least* different from the Sample. Palamedes provides eight routines for the 2AFC match-to-sample task:

```
PAL_SDT_2AFCmatchSample_DiffMod_PCToDP  
PAL_SDT_2AFCmatchSample_DiffMod_DPToPC  
PAL_SDT_2AFCmatchSample_DiffMod_PHFtoDP  
PAL_SDT_2AFCmatchSample_DiffMod_DPToPHF  
PAL_SDT_2AFCmatchSample_IndMod_PCToDP  
PAL_SDT_2AFCmatchSample_IndMod_DPToPC
```

PAL_SDT_2AFCmatchSample_IndMod_PHFtoDP
PAL_SDT_2AFCmatchSample_indMod_DPtоНР

The routines use the same input and output arguments as the same-different routines. Given that observers might be biased towards choosing one Match alternative over the other, it is recommended to use the routines that take pH and pF rather than Pc as arguments, unless there is good reason to assume the observer is unbiased.

6.2.6.2 d' for M-AFC Match-to-Sample

For $M > 2$ match-to-sample tasks, Palamedes has two routines:

PAL_SDT_MAFMatchSample_DPtоС and
PAL_SDT_MAFMatchSample_PCtoDP.

Both routines assume that the observer is unbiased and adopts a differencing strategy. In keeping with other Palamedes SDT routines, each routine takes two input arguments: a scalar, vector or matrix of Pcs or d' s, and a value of M . The output arguments are d' s or Pcs . The reader will find that the **_DPtoPC** routine is slower to execute than other SDT routines. This is because the calculations are implemented by Monte Carlo simulation using a very large number of trials. The reverse routine, **_PctoDP**, is even slower as it performs an iterative search based on the forward routine. As a result, the routine may take minutes to execute depending on the speed of the computer and the number of input Pc values.

6.2.7 Measuring d' for M-AFC Oddity Tasks

In an oddity task, often termed an "odd-man-out" task, the observer is presented with an array of stimuli, all but one of which are the same, and chooses the stimulus that is different, in other words the "oddity." As with the same-different and match-to-sample tasks, the observer in an oddity task does not need to know the basis on which the stimuli differ. Probably the most popular form of oddity task is the one using the minimum number of alternatives per trial, which is three, and for this reason sometimes termed the "triangular" method. However, the principle extends to any M . One likely strategy in an oddity task is that observers select the alternative that is most different from the mean of all the alternatives, another instance of a differencing strategy. Craven (1992) has provided a table for converting Pc into d' for oddity tasks with various M , assuming this strategy and an unbiased observer. Palamedes provides two routines that perform the same computations as those described in Craven (1992), but for any M : **PAL_SDT_MAFOddity_DPtоС** and **PAL_SDT_MAFOddity_PCtoDP**. As elsewhere, each routine takes two arguments: a scalar, vector or matrix of Pcs or d' s, and the value of M . The output arguments are d' s or Pcs . As with the M-AFC match-to-sample routines described in

the previous section, the M -AFC oddity routines are slow to execute as they also employ Monte Carlo simulation.

6.2.8 Estimating $P_{c_{\max}}$ with Observer Bias

As we have argued above, P_c is not a valid measure of performance for any of the procedures described if there is a significant amount of observer bias. However, it is possible to obtain an estimate of the P_c that would be expected if the observer were not biased. This is termed $P_{c_{\max}}$ (or $P_{c_{\text{unb}}}$), because P_c reaches a theoretical maximum when there is no bias (e.g., see Figure 6.2b for the 1AFC task). One can think of $P_{c_{\max}}$ as an unbiased estimate of P_c . Estimating $P_{c_{\max}}$ is straightforward with Palamedes, provided one has available a measure of the criterion that is zero when the observer is unbiased, as with the routines that compute the criterion measure C . To obtain $P_{c_{\max}}$ one inputs pH and pF into the relevant routine (i.e., one ending in **_PHFToDP**) to obtain d' and a measure of C , and then use the reverse routine (the same routine ending in **_DPToPHF**) to convert back to pH and pF , using as the input argument a zero value for C . $P_{c_{\max}}$ is then equal to the output pH .

Take the following example. Suppose you want to estimate $P_{c_{\max}}$ for a 2AFC match-to-sample task assuming a differencing strategy. Let $pH = 0.8$ and $pF = 0.6$. One can glean from these values that the observer is biased towards the alternative for which a correct response is classified as a "hit," since the number of false alarms exceeds $1 - 0.8$ i.e. 0.2. Recall also that P_c is given by $[pH + (1 - pF)]/2$, which for this example is 0.6. If the values of pH and pF are input to **PAL_SDT_2AFCmatchSample_DiffMod_PHFToDP**, the routine returns a d' of 1.2137 and a criterion C of -0.5475. If one now inputs the same d' to **PAL_SDT_2AFCmatchSample_DiffMod_PHFToDP**, but with C set to zero, the outputs are $pH = 0.6157$ and $pF = 0.3843$. Thus, $P_{c_{\max}}$ is 0.6157. $P_{c_{\max}}$ is only slightly higher than the actual P_c because the bias in the example is not particularly strong.

6.2.9 Comparing d' 's and P_c s across Different Tasks

Two scripts are provided by Palamedes that demonstrate the differences between the computed d' 's and P_c 's for a variety of tasks: **PAL_SDT_DPtоСCcomparison_Demo** and **PAL_SDT_PCToDPcomparison_Demo**. The tasks compared are 1AFC, standard 2AFC, 2AFC same-different, and 2AFC match-to-sample. The standard 2AFC, same-different, and match-to-sample tasks assume a differencing strategy, and all tasks assume an unbiased observer. Therefore, for the 1AFC tasks, criterion C is set to zero to produce an optimal P_c . The scripts prompt you either for d' 's or P_c 's. Try the first program:

```
>>PAL_SDT_DPtоСCcomparison_Demo
```

Enter a vector of **Dprime** values and enter:

```
[0:.5:4]
```

The output should look like this:

-----Proportion correct-----

dprime	1AFC	2AFC	2AFCsameDiff	2AFCmatchSamp
0	0.5000	0.5000	0.5000	0.5000
0.5000	0.5987	0.6382	0.5195	0.5223
1.0000	0.6915	0.7602	0.5733	0.5825
1.5000	0.7734	0.8556	0.6495	0.6635
2.0000	0.8413	0.9214	0.7330	0.7468
2.5000	0.8944	0.9615	0.8110	0.8196
3.0000	0.9332	0.9831	0.8753	0.8765
3.5000	0.9599	0.9933	0.9231	0.9178
4.0000	0.9772	0.9977	0.9555	0.9467

and a graph will be plotted as in Figure 6.3 shown above.

6.3 SECTION B: THEORY

6.3.1 Relationship Between Z-scores and Probabilities

To understand the theory behind calculations of d' it is necessary to begin with some basics. An important relationship that underpins much of SDT is that between z -values and probabilities. Figure 6.4 shows a "standardized" normal probability distribution. This is a normal distribution in which the abscissa is given in units of standard deviation, or z units. The ordinate in the graph is termed "probability density" and denoted by ϕ . Probability density values are not actual probabilities of z -values, but their relative likelihoods, specifically derivatives or "rates of change" of probabilities. Thus, in order to convert z units, or rather intervals between z units, into probabilities, one has to integrate the values under the curve between z -values. If one integrates the curve between $-\infty$ and some value of z , the result is a value from a distribution termed the cumulative normal. Because the total area under the standard normal distribution is by definition unity, the cumulative normal distribution ranges from 0–1. The cumulative normal gives the probability that a random variable from a standardized normal distribution is less than or equal to z .

The equation for the standardized normal distribution is:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) \quad (6.1)$$

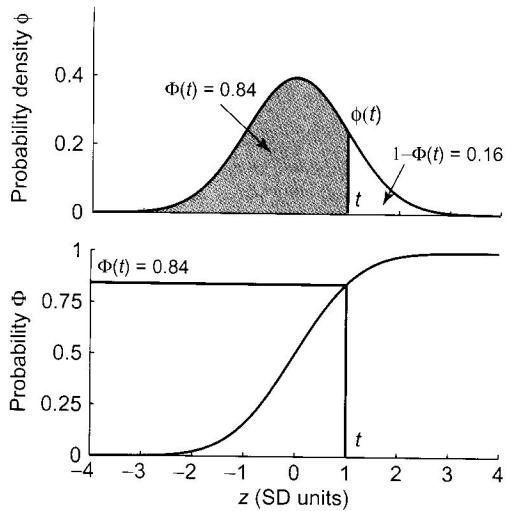


FIGURE 6.4 Relationship between z , probability density ϕ , and cumulative probability Φ . Top: standardized normal distribution. Bottom: the integral of the standardized normal distribution, or cumulative probability distribution. The value of z at point t is 1 in both graphs. In the top graph the height of the distribution at t is denoted by $\phi(t)$ and the area under the curve to the left of t (shown in gray) which has a value of 0.84 denoted as $\Phi(t)$. In the bottom graph $\Phi(t)$ is now a point on the ordinate. In the top graph the white area to the right of t , defined as $1 - \Phi(t)$ has a value of 0.16.

and for the cumulative normal:

$$\Phi(z) = 0.5 + 0.5\text{erf}(z/\sqrt{2}) \quad (6.2)$$

where erf stands for the “error function”, which performs the integration. The two values of 0.5 in the equation convert the range of the function to 0–1. The inverse of the cumulative normal, which converts Φ to a z -value is:

$$z(\Phi) = \sqrt{2}\text{erfinv}(2\Phi - 1) \quad (6.3)$$

Palamedes contains two routines, **PAL_ZtoP** and **PAL_PtoZ**, which implement, respectively, Equations 6.2 and 6.3. Given that z -values are symmetric around zero, we can state two simple relationships:

$$1 - \Phi(z) = \Phi(-z) \quad (6.4)$$

and

$$-z(\Phi) = z(1 - \Phi) \quad (6.5)$$

You can verify these relationships and the values illustrated in the figure using **PAL_ZtoP** and **PAL_PtoZ**.

6.3.2 Calculation of d' for M-AFC

We begin by describing the theory behind the computation of d' for a standard M-AFC task, where M can be any value greater than 1, and where $M = N = m$. We remind the reader that the calculations described in this section are based on two assumptions, the first that the observer is unbiased and the second that the internal responses to all the stimulus alternatives are normally distributed and of equal variance. Although readers will mainly use the routine **PAL_SDT_MAFC_PCToDP**, which converts P_c s to d' s, it is best to begin with the theory behind its inverse: **PAL_SDT_MAFC_PCToDP**.

Figure 6.5 shows two standardized normal distributions. One represents the distribution of sensory magnitudes or internal responses to a "blank" interval or location, i.e., one without a target and denoted by "noise alone" or N . The other represents the distribution of sensory magnitudes to the interval/location containing the target, typically denoted in the SDT literature as "signal-plus-noise" or $S + N$. Note, however, that N versus $S + N$ is not the only scenario for which the present analysis is applicable. The principle also extends to the situation in which one interval/location contains stimulus S_1 while the remaining intervals/locations contain stimulus S_2 .

Representing the sensory magnitudes of N and $S + N$ as probability distributions means that on any trial the actual sensory magnitudes will be random samples from those distributions. The relative probabilities of particular samples are given by the heights of the distributions at the sample points.

The aim of the observer in the standard forced-choice task is to identify on each trial the alternative containing the target. Let us assume that the observer adopts what is intuitively the optimum strategy: select the alternative with the biggest signal. Try to imagine a strategy that would result in better performance. There isn't one. The rule employed by the observer for selecting the target is usually termed the "decision rule." The question then becomes: how well will the observer do, as measured by P_c ,

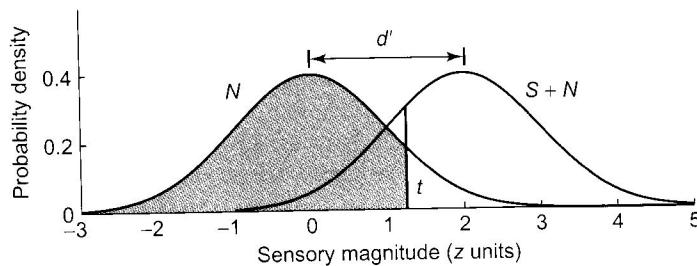


FIGURE 6.5 Calculation of P_c from d' . N = noise; $S + N$ = signal-plus-noise; t is a random variable. See text for details.

when adopting this decision rule? If we make the two assumptions stated above, then the computation of d' turns out to be reasonably straightforward.

One can glean from Figure 6.5 that when there is little overlap between the N and $S + N$ distributions the observer will perform better than when there is a lot of overlap. The reason for this is that, as the N and $S + N$ distributions draw closer together, there is an increasing likelihood that a sample drawn randomly from the N distribution will be greater in magnitude than a sample drawn randomly from the $S + N$ distribution. Each time this happens the observer will make a mistake if adopting the "select the biggest signal" decision rule. If there were no overlap at all between the two distributions, the observer would never make an incorrect decision using this rule, and if the distributions perfectly overlapped the observer would perform at chance. Thus, the degree of overlap between the two distributions is the critical determinant of performance. And because the overlap is governed by two factors, first the separation of the two distributions and second their spread, or σ , one can see that the measure d' , which is the separation between the distributions expressed in units of σ , captures the discriminability of N and $S + N$. But how do we calculate the expected P_c , given d' and M ?

Suppose that on a given trial the target stimulus has a sensory magnitude given by t in the figure. Remember that t is a random sample, meaning that t will vary between trials, and that the relative probability of a given t is given by the height of the distribution at t . The probability that t will be greater than a random sample from just one noise (N) location is given by the gray area to the left of t under the noise distribution. This is simply $\Phi(t)$, since we have (arbitrarily) centered the noise distribution at zero. However, we do not just wish to know the probability that t will be greater than a random sample from just one noise location, but from that t will be greater than a random sample from just one noise location, but from $M - 1$ noise locations. In other words, we want to know the probability that t will be greater than a random sample from noise location 1 and noise location 2 and be greater than a random sample from noise location 3 and 4 and so on, up to $M - 1$. The "and" term here implies a joint probability, and if we assume that the samples from the different noise locations are independent, this is obtained simply by multiplying the individual probabilities. Since we are multiplying the same thing over again we simply raise the probability to the power of $M - 1$, and hence obtain $\Phi(t)^{M-1}$. However, this still only gives us the probability that one specific random sample from the signal distribution, t , will be greater than all random samples from all $M - 1$ noise locations. To obtain the probability that a random sample t will be greater than random samples from $M - 1$ noise locations, which gives us our P_c , we need to integrate the above result across all possible values of t . We do this by multiplying $\Phi(t)^{M-1}$ by the height, or relative likelihood of t , which is given by $\phi(t - d')$ (the S distribution is offset from zero by d'), and integrating over all possible values of t . Hence we have:

$$P_c = \int_{-\infty}^{\infty} \phi(t - d') \cdot \Phi(t)^{M-1} dt \quad (6.6)$$

(Green & Swets, 1974; Wickens, 2002). The function **PAL_SDT_MAFC_DPttoPC** implements this equation using the numerical integration function **quadgk** in MATLAB®.

How do we convert a P_c into a d' for an $M - \text{AFC}$ task, which is our primary aim? Equation 6.6 is not easily invertible, so **PAL_SDT_MAFC_PCToDP** performs an iterative search using the **fminsearch** function in MATLAB to find that value of d' which, when converted to P_c (using **PAL_SDT_MAFC_DPttoPC**), gives the input P_c .

6.3.3 Calculation of d' and Measures of Bias for 1AFC Tasks

6.3.3.1 Calculation of d' for 1AFC

Let us consider the 1AFC task known as yes/no, a task that is particularly prone to bias. Adopting the same scheme for representing the distributions of sensory magnitudes as in the previous section for the standard M -AFC task, the situation is illustrated in Figure 6.6. This time, the N and $S + N$ distributions are shown separately as the stimuli they represent are presented on separate trials. The gray areas to the right of the vertical criterion line represent sensory magnitudes that the observer deems large enough to warrant a "yes" response. Sensory magnitudes to the left of this line produce a "no" response. The gray area to the right of the criterion in the lower $S + N$ distribution gives the proportion of target-present

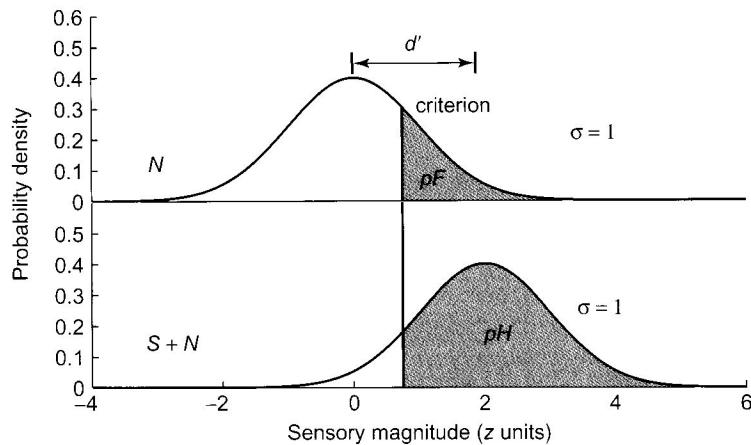


FIGURE 6.6 Distributions of sensory magnitude in response to both noise N and signal-plus-noise ($S + N$) in a 1AFC yes/no task. The vertical black line shows the position of the observer's criterion. Sensory magnitudes to the right of this line result in a "yes" response, while those to the left a "no" response. pH is the proportion of "hits," or correct "yes" responses, and pF the proportion of "false alarms," i.e., incorrect "yes" responses. pH and pF are given by the gray areas to the right of the criterion line.

trials resulting in a "yes" response, i.e., the proportion of hits or pH . The gray area to the right of the criterion line in the upper N distribution gives the number of "yes" responses in target-absent trials, i.e., the proportion of false alarms or pF . If we denote the position of the criterion line on the abscissa as c (see below), then one can see that:

$$pF = 1 - \Phi(c)$$

$$\text{or } pF = \Phi(-c)$$

and

$$pH = 1 - \Phi(c - d')$$

$$\text{or } pH = \Phi(-c + d')$$

Converting pF and pH to z -values one obtains:

$$z(pF) = -c$$

and

$$z(pH) = -c + d'$$

Combining these two equations and solving for d' gives

$$d' = z(pH) - z(pF) \quad (6.7)$$

6.3.3.2 Calculation of Criterion C for 1AFC

In Figure 6.6 it can be seen that the criterion is measurable in z units, with a high z -value implying a strict criterion (few hits but few false alarms), and a low z -value implying a loose criterion (many hits but many false alarms). However, the actual criterion z -value depends on where the zero z -value is positioned, so a convention is needed to ensure that the criterion measure is comparable across conditions. The convention is to place the zero point midway between the N and $S + N$ distributions, as shown in Figure 6.7.

With $z = 0$ centered midway between the two distributions, the criterion, denoted by C , is positioned in the noise distribution at:

$$z(1 - pF) - d'/2$$

and in the signal-plus-noise distribution at:

$$z(1 - pH) + d'/2$$

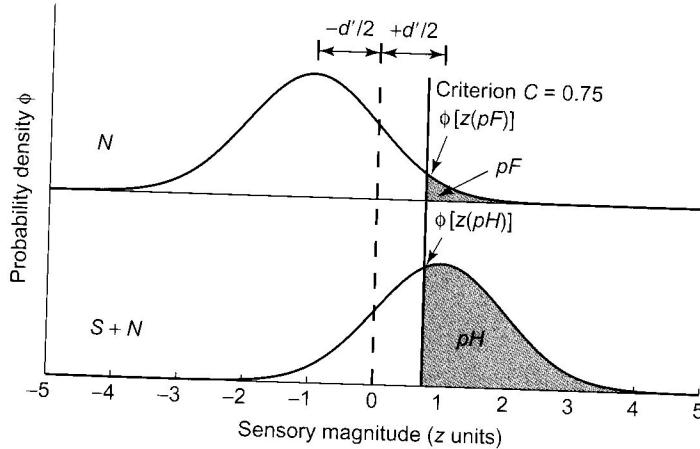


FIGURE 6.7 Method for calculating criterion C . Note that $z = 0$ is centred midway between the N and $S + N$ distributions. See text for further details.

However, since $z(p) = z(1 - p)$, the two expressions can be rewritten as $-z(pF) - d'/2$ and $-z(pH) + d'/2$. Thus, the position of C can be defined in two ways:

$$C = -z(pH) + d'/2$$

and

$$C = -z(pF) - d'/2$$

Adding the two equations together gives:

$$C = -[z(pH) + z(pF)]/2 \quad (6.8)$$

(Macmillan & Creelman, 2005). Thus, criterion C can be estimated by converting pH and pF into z -values, and then using Equation 6.8. C is calculated this way in **PAL_SDT_1AFC_PHFToDP**. C can range from negative to positive, with negative values indicating a bias towards "yes" responses and positive values a bias towards "no" responses.

6.3.3.3 Calculation of Criterion $\ln\beta$ for 1AFC

An alternative measure of the criterion is the natural logarithm of the ratio of the heights of the two distributions at C (Macmillan & Creelman, 2005). The heights at C are shown in Figure 6.7 as $\phi[z(pH)]$ and $\phi[z(pF)]$. Thus:

$$\ln\beta = \ln \frac{\phi[z(pH)]}{\phi[z(pF)]} \quad (6.9)$$

Now $\phi[z(pH)]$ and $\phi[z(pF)]$ are given by:

$$\phi[z(pH)] = \frac{1}{\sqrt{2\pi}} \exp\left[\frac{-\{-z(pH)^2\}}{2}\right] \quad (6.10)$$

and

$$\phi[z(pF)] = \frac{1}{\sqrt{2\pi}} \exp\left[\frac{-\{-z(pF)^2\}}{2}\right] \quad (6.11)$$

Taking natural logarithms of the two equations, i.e., $\ln\{\phi[z(pH)]\}$ and $\ln\{\phi[z(pF)]\}$ and then substituting the results into the equation for $\ln\beta$, simple algebra shows that:

$$\ln\beta = [z(pF)^2 - z(pH)^2]/2 \quad (6.12)$$

This is how $\ln\beta$ is calculated in **PAL_SDT_1AFC_PHFtoDP**. $\ln\beta$ behaves in the same way as C . The reason for this is that $\ln\beta = Cd'$. Readers can check this relationship themselves using the equations above.

6.3.3.4 Calculation of Criterion C' for 1AFC

A third measure of the criterion not calculated by Palamedes is C' , which is C expressed as a proportion of d' (Macmillan & Creelman, 2005):

$$C' = C/d' = \frac{-[z(pH) + z(pF)]}{2[z(pH) - z(pF)]} \quad (6.13)$$

6.3.3.5 Calculation of $P_{C_{max}}$ for 1AFC

In Section A we showed graphically that with a 1AFC task the optimum P_C , or $P_{C_{max}}$, is obtained when the observer is unbiased, i.e., when $C = 0$. It follows from Equation 6.8 that when $C = 0$, $z(pH) = -z(pF)$. Since $d' = z(pH) - z(pF)$ (Equation 6.7), simple algebra reveals that when $C = 0$, $d' = 2z(pH)$. Converting $z(pH)$ to $P_{C_{max}}$ gives:

$$P_{C_{max}} = \Phi(d'/2) \quad (6.14)$$

The interested reader may wish to prove that P_C reaches a maximum when $C = 0$. One can determine if the observer is operating optimally in a 1AFC task by testing whether $pH = 1 - pF$. When performing optimally, pH is $P_{C_{max}}$ and d' can be calculated as $2z(pH)$.

6.3.4 Calculation of d' for Unbiased and Biased 2AFC Tasks

In the first section of Section B we derived the formula for calculating d' from P_c for an unbiased standard M -AFC task (Equation 6.6). This formula can be used to calculate d' for the standard 2AFC task ($M = 2$), assuming that the observer is unbiased. In the following sections we show a simpler method for calculating d' for an unbiased 2AFC task, and show how d' and measures of bias can be calculated for 2AFC tasks in which the observer is biased.

6.3.4.1 Alternative Calculation of d' for Unbiased 2AFC

With the standard 2AFC procedure, the N and $S + N$ stimuli are presented together in a trial as two alternatives. Remember that the decision rule is to choose the alternative in which the internal signal is biggest. If the observer adopts this rule, trials in which the *differences* between the $S + N$ and N samples are positive will result in a correct decision. Now the distribution of differences between random samples from two equal-variance normal distributions, one with mean 0 the other with mean d' , is a normal distribution with a mean of d' and a variance of 2, i.e., a σ of $\sqrt{2}$. The σ of $\sqrt{2}$ follows from the variance sum law. This law states that the variance of the sum, or of the difference, between two uncorrelated random variables is the sum of the variances of the two variables. Thus, if the two distributions each have a σ of 1, the σ of the difference between the two distributions is $\sqrt{(1^2 + 1^2)} = \sqrt{2}$. The $(S + N) - N$ difference distribution is illustrated in the lower panel of Figure 6.8. Note that in this graph the abscissa is in z units that have been normalized to the σ s of the N and $S + N$ distributions, not to the σ of their difference.

The proportion correct for 2AFC is thus given by the gray area in the lower panel to the right of zero. This is:

$$P_c = \Phi\left(\frac{d'}{\sqrt{2}}\right) \quad (6.15)$$

(Wickens, 2002; Macmillan & Creelman, 2005; McNicol, 2002). Equation 6.15 converts a z -value of $d'/\sqrt{2}$ to a Φ value. One can implement the equation using **PAL_ZtoP** using as an input argument a scalar, vector or matrix of $d'/\sqrt{2}$. However, in most instances we want to obtain d' from P_c , so for this we use the inverse equation:

$$d' = z(P_c)\sqrt{2} \quad (6.16)$$

The following converts a vector **PropCorr** containing an array of P_c s into a vector **DP** containing an array of d' s, for an unbiased 2AFC task:

```
>> DP=PAL_PtoZ(PropCorr)*sqrt(2)
```

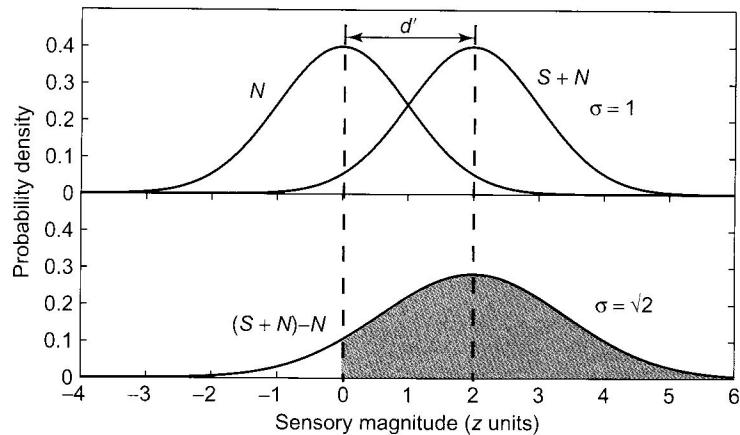


FIGURE 6.8 Graphical illustration of how d' can be calculated for an unbiased 2AFC task. Top: distributions of noise alone (N) and signal-plus-noise ($S + N$) separated by d' . Bottom: distribution of the difference between the two distributions: $(S + N) - N$. Note the different σ s for the upper and lower distributions. The z values along the abscissa are normalized to the σ of the two distributions in the top, not bottom panel. See text for further details.

6.3.4.2 Calculation of d' for Biased 2AFC

Let us consider the situation in which the two alternatives are presented sequentially, i.e., 2IFC. Figure 6.9 plots the distribution of differences in sensory magnitude between those in the first interval (X_1) and those in the second interval (X_2), i.e., the distribution of $X_1 - X_2$. Note that there are now two distributions, one for signal present in the first interval and one for signal present in the second interval. The two distributions will be separated by $2d'$ and have σ s of $\sqrt{2}$ (see above). If the observer is biased towards responding to one interval more than the other, then their criterion C will be non-zero. The observer's decision rule is "1" (first interval) if $X_1 - X_2 > C$, and "2" (second interval) if $X_1 - X_2 < C$. As explained in Section A, the key to calculating d' for a biased 2AFC task is to classify the responses in terms of hits and false alarms, where a "1" response is scored as a hit when the signal is in the first interval and a false alarm when the signal is in the second interval.

One can see from Figure 6.9 that:

$$z(pH) = (d' - C)/\sqrt{2} \quad (6.17)$$

and

$$z(pF) = (-d' - C)/\sqrt{2} \quad (6.18)$$

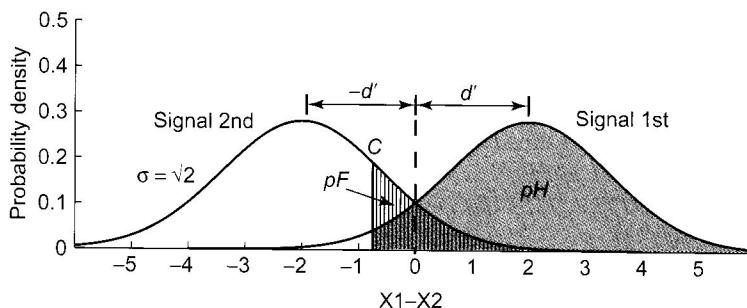


FIGURE 6.9 Relationship between d' , C , pH , and pF in a biased 2AFC task. Each plot gives the distribution of differences between the sensory magnitudes in the first (X_1) and second (X_2) alternatives/intervals. If the signal is in the first alternative/interval the distribution is the one shown on the right, if in the second interval the distribution shown on the left.

Combining the two equations and solving for d' gives:

$$d' = [z(pH) - z(pF)] / \sqrt{2} \quad (6.19)$$

This is the equation used to calculate d' in **PAL_SDT_2AFC_PHFtoDP**.

6.3.4.3 Calculation of C and $\ln\beta$ for Biased 2AFC

Combining Equations 6.17 and 6.18 above and solving for C gives:

$$C = -[z(pH) + z(pF)] / \sqrt{2} \quad (6.20)$$

The criterion measure $\ln\beta$ is defined in Equation 6.9, and it is important to note that the ϕ s refer to the heights of $z(pH)$ and $z(pF)$ in the standard normal distribution, i.e., the distributions in the upper panel of Figure 6.8, not to the heights of the difference distributions in Figure 6.9. The calculation of $\ln\beta$ for the 2AFC task is thus identical to that for the 1AFC task (we would like to thank Mark Georgeson for pointing this out), and is hence given by Equation 6.12. Equation 6.20 is used to calculate bias for the standard 2AFC task by **PAL_SDT_2AFC_PHFtoDP**.

6.3.4.4 Calculation of $P_{c_{max}}$ for 2AFC

From Equation 6.20, if $C = 0$, then $z(pH) = -z(pF)$. Combining this result with Equation 6.19 reveals that when $C = 0$, $d'\sqrt{2} = 2z(pH)$. Converting $z(pH)$ to $P_{c_{max}}$ gives:

$$P_{c_{max}} = \Phi(d'/\sqrt{2}) \quad (6.21)$$

6.3.5 Calculation of d' for Same-Different Tasks

For the calculation of d' for a same-different task we adopt the convention of referring to the two relevant distributions as S_1 and S_2 (signal 1 and 2), rather than N and $S + N$. It would be unusual to employ a same-different task to measure the detectability of a target when the alternative location/interval was a blank. The same-different task is most appropriate to situations in which the observer is required to discriminate two suprathreshold stimuli without necessarily having to know the basis of the discrimination.

6.3.5.1 Calculation of d' for a 2AFC Same-Different

The computation of d' for the same-different task in which both the same *and* different pairs are presented together during a trial is described by Macmillan, Kaplan, and Creelman (1977). They use the term 4IAX to characterize the task, since they consider the scenario in which the four stimuli are presented in temporal order, as in an auditory experiment.

Let us begin with the standard assumption that the sensory magnitudes of S_1 and S_2 are normally distributed and separated by d' . According to Macmillan et al. (1977), the most likely strategy employed by observers in this task is to compare the absolute difference between the two signals in each of the first and second pairs. The observer responds "1" if the difference between the first pair is perceived to be greater than the difference between the second pair, and "2" otherwise. Suppose that the sensory magnitudes of the four stimuli are represented by the sequence X_1, X_2, X_3 , and X_4 . The decision rule is therefore to respond "1" if $|X_1 - X_2| > |X_3 - X_4|$, and "2" if $|X_1 - X_2| < |X_3 - X_4|$.

Figure 6.10, adapted from Macmillan et al. (1977), illustrates the computation of d' for the task. The abscissa and ordinate in Figure 6.9 represent, respectively, the decision variables $X_1 - X_2$ and $X_3 - X_4$. The gray areas in the figure represent the combinations of decision variables that result in a "1" decision, i.e., areas where $|X_1 - X_2| > |X_3 - X_4|$. The gray areas can be subdivided into four regions: upper left; lower left; upper right; and lower right. On the right side of the figure the gray area defines the space in which $X_1 - X_2$ is more positive than either $X_3 - X_4$ (upper right) or $-(X_3 - X_4)$ (lower right). On the left of the figure the gray area defines the space in which $X_1 - X_2$ is more negative than either $X_3 - X_4$ (lower left) or $-(X_3 - X_4)$ (upper left).

The observer will be correct when making a "1" decision if the samples that fall within the gray regions are from any of the following sequences: $\langle S_1 S_2 S_1 S_1 \rangle$, $\langle S_1 S_2 S_2 S_2 \rangle$, $\langle S_2 S_1 S_1 S_1 \rangle$ or $\langle S_2 S_1 S_2 S_2 \rangle$. On the other hand, the observer will be incorrect when responding "1" if samples from the remaining sequences fall within the gray area, namely $\langle S_1 S_1 S_1 S_2 \rangle$, $\langle S_1 S_1 S_2 S_1 \rangle$, $\langle S_2 S_2 S_1 S_2 \rangle$ or $\langle S_2 S_2 S_2 S_1 \rangle$. P_c is therefore the probability that samples from the first four sequences will fall within either of the two (left or right) gray areas. The four rings in the figure denote

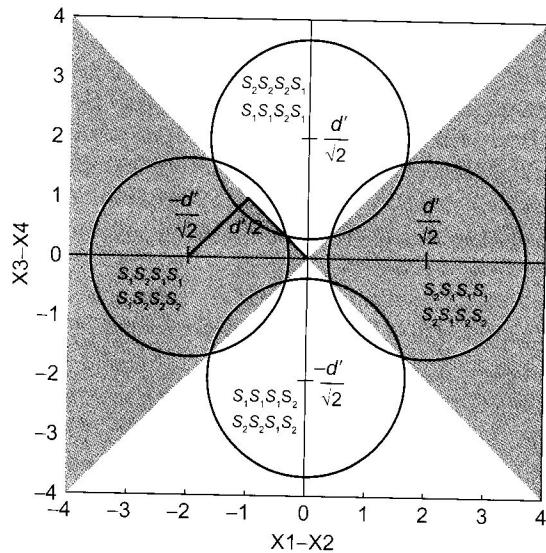


FIGURE 6.10 Graphical representation of the distributions involved in the 2AFC same-different task. $X_1 \dots X_4$ represent the internal sensory magnitudes of the four stimuli. Note that the abscissa plots $X_1 - X_2$ and the ordinate $X_3 - X_4$. The sequences $\langle S_1 S_2 S_1 S_1 \rangle$ etc., denote joint sample distributions of stimulus sequences. Note that the distance to the center of each distribution from the center of the figure is $d' / \sqrt{2}$, but when measured from a point on the diagonal perpendicular to the center of each distribution (shown by the thick black lines in the upper left quadrant) the distance is $d' / 2$. The figure is adapted from Figure 6a in Macmillan et al., (1977).

volumes of the joint likelihood distributions of the various sequences of S_1 and S_2 . Note that the σ of the distributions is $\sqrt{2}$, because they are distributions of the difference between samples from two normal distributions.

Each volume in the left and right gray areas comprises two probabilities, a "small" and a "large." The large probability is the probability that samples from the sequences specified within each gray area of the figure will fall within that area. However, there is a small probability that samples from the sequences in the opposite gray area will also fall within the area. For example, although most of the samples that fall within the gray area on the right of the figure will come from sequences $\langle S_2 S_1 S_1 S_1 \rangle$ and $\langle S_2 S_1 S_2 S_2 \rangle$, a few will come from $\langle S_1, S_2, S_1, S_1 \rangle$ and $\langle S_1, S_2, S_2, S_2 \rangle$. This is because even though most of the difference signals $S_1 - S_2$ are "large negative" and hence fall within the gray area on the left, a few will be "large positive" and will fall within the gray area on the right. Remember that it does not matter whether the difference $S_1 - S_2$ is "large negative" or "large positive," as long as its absolute magnitude is greater than $S_2 - S_2$ or $S_1 - S_1$ (the possible sequences in the other alternative/interval). Either way the response "1" will be correct.

The larger probability within each gray area is given by $[\Phi(d'/2)]^2$ while the smaller probability is given by $[\Phi(-d'/2)]^2$. The denominator of 2 in each expression reflects the fact that the area described by the gray rectangle has sides that, by the Pythagorean Theorem, extend by $d'/2$ to the midpoint of the distribution along the side, as illustrated in the upper left quadrant of the figure. The squaring of each expression reflects the fact that one is dealing with a bivariate, i.e., joint, distribution. To obtain P_c we simply add together the large and small probabilities:

$$P_c = [\Phi(d'/2)]^2 + [\Phi(-d'/2)]^2 \quad (6.22)$$

Equation 6.22 is used to calculate P_c in **PAL_SDT_2AFCsameDiff_DPttoPC**. Following Macmillan and Creelman (2005), the equation can be inverted to obtain d' from P_c using:

$$d' = 2z[0.5\{1 + (2P_c - 1)^2\}] \quad (6.23)$$

and this is used to calculate d' in **PAL_SDT_2AFCsameDiff_PCToDP**.

6.3.5.2 Calculation of d' for a 1AFC Same-Different Task: Differencing Model

In the differencing model of the 1AFC same-different task it is assumed that the observer encodes the perceived difference between the two stimuli in the trial, and if the absolute value of the difference exceeds a criterion the observer responds "different," if not "same." Suppose the signal from the first stimulus is X_1 , and from the second stimulus X_2 . The decision rule is therefore "different" if $|X_1 - X_2| > k$, where k = the criterion, and "same" otherwise. As with the 2AFC same-different task discussed in the previous section, it is useful to consider both the positive and negative parts of the difference signal $X_1 - X_2$. The top of Figure 6.11 shows the distributions of sensory magnitudes for the two stimuli S_1 and S_2 , centered on 0. The middle and bottom panels (adapted from Figure 9.5 in Macmillan & Creelman, 2005) show the relative likelihoods of the various stimulus pairs as a function of the decision variable $X_1 - X_2$. The middle panel shows the Same distributions $\langle S_1 S_1 \rangle$ and $\langle S_2 S_2 \rangle$, and the bottom panel the Different distributions $\langle S_1 S_2 \rangle$ and $\langle S_2 S_1 \rangle$.

All the Different distributions have a σ of $\sqrt{2}$, in accordance with the variance sum law. To understand how pH and pF are calculated the criterion has been placed to one side of the midpoint. Given the particular value of d' and k in the figure, most of the $\langle S_2 S_1 \rangle$ signals fall above the criterion k and constitute a "large" probability. Although most of the $\langle S_1 S_2 \rangle$ signals fall to the left of k , a few will be "large positive" and fall to its right. As with the 2AFC same-different task we have to include the small probability in the calculation, because it accords with the adopted decision rule. From Figure 6.11 the proportion of hits, pH is given by:

$$pH = \Phi[(d' - k)/\sqrt{2}] + \Phi[(-d' - k)/\sqrt{2}] \quad (6.24)$$

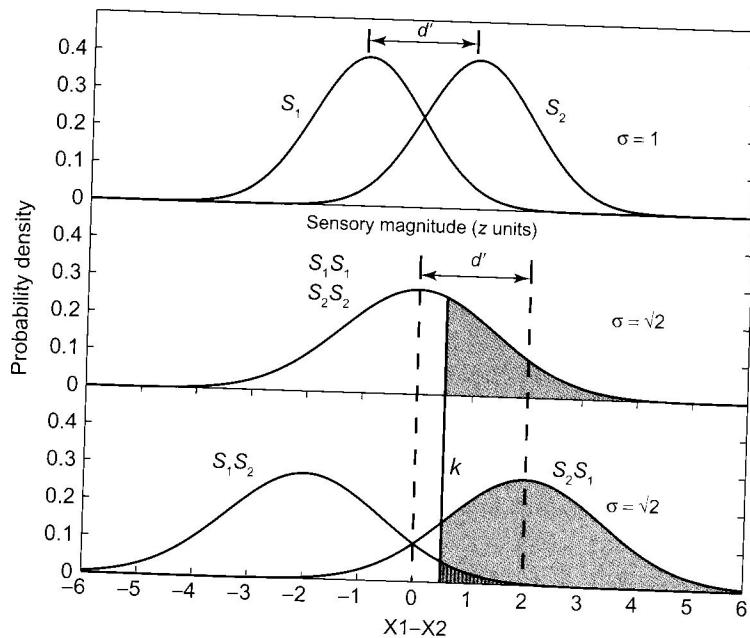


FIGURE 6.11 Method for calculating d' for a 1AFC same-different task assuming a differencing model. See text for details.

where the larger of the two terms is given by the gray area to the right of k and the smaller of the two terms by the hatched area to the right of k . The proportion of false alarms pF is given by the area to the right of the criterion line in the middle panel, multiplied by 2 since there are two distributions, i.e.:

$$pF = 2\Phi(-k/\sqrt{2}) \quad (6.25)$$

The routine **PAL_SDT_1AFCsameDiff_DiffMod_DPttoPHF** performs these calculations. To calculate d' and k from pH and pF , as is mostly required, the routine **PAL_SDT_1AFCsameDiff_DiffMod_PHFtoDP** exploits the fact that k can be obtained directly from pF , as from Equation 6.24 $k = -z(pF/2)\sqrt{2}$. The value of k is then substituted into Equation 6.24 and the routine performs an iterative search to find that value of d' that results in the input value of pH . Further details of the 1AFC same-different differencing model can be found in Macmillan and Creelman (2005).

6.3.5.3 Calculation of d' for a 1AFC Same-Different Task: Independent Observation Model

According to Macmillan and Creelman (2005), the observer's optimum strategy for the same-different task in which only two stimuli are presented per trial

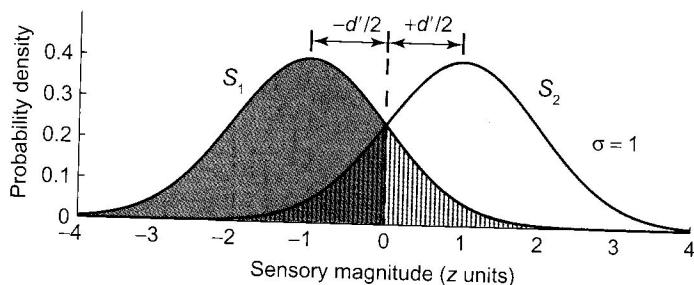


FIGURE 6.12 Principle behind the computation of d' for the independent observation model of the 1AFC same-different task. See text for details.

is to respond "different" when the signals from S_1 and S_2 fall on either side of a criterion centered midway between the two distributions. They term this model the independent observation model. To compute d' for this model, Macmillan and Creelman suggest the following method. First, calculate the P_c that an observer would obtain for this task if they were operating optimally (this is $P_{c_{\max}}$) using p_H and p_F . Second, use $P_{c_{\max}}$ to compute d' . Third, use the values of p_H and p_F to compute the criterion C in the same way as for the standard 1AFC task.

As elsewhere, it is best to begin with the method for calculating $P_{c_{\max}}$ from d' , rather than the reverse. Macmillan and Creelman (2005) provide a three-dimensional representation of the decision space for the independent observation model, as the calculations involve joint likelihood distributions. The two-dimensional representation provided in Figure 6.12 should, however, be sufficient to understand the principle behind the calculation.

In Figure 6.12, the probability that signals from both S_1 and S_2 will fall on opposite sides of the criterion at zero is the probability that S_1 falls to the left of the criterion multiplied by the probability that S_2 falls to its right (since we are dealing here with the joint probability of an event). In the figure, given the value of $d' = 2$, most of the S_2 signals fall to the right of the criterion and most of the S_1 signals will fall to the left of the criterion, so the product of the two signals will be a "large" probability given by $[\Phi(d'/2)]^2$. However, there is a small probability that both a high value of S_1 and a low value of S_2 will fall on either side of the criterion. These probabilities are the smaller hatched areas in the figure. The observer will also be correct in these instances, since the decision rule is to respond "different" when the signals from the two stimuli fall on either side of the criterion. The joint probability in this case is given by the product of the hatched areas, which is $[\Phi(-d'/2)]^2$. Thus, to obtain $P_{c_{\max}}$ we add up the two joint probabilities:

$$P_{c_{\max}} = [\Phi(d'/2)]^2 + [\Phi(-d'/2)]^2 \quad (6.26)$$

and from this equation, d' is given by:

$$d' = 2z\{0.5[1 + \sqrt{2P_{c_{\max}} - 1}]\} \quad (6.27)$$

(Macmillan & Creelman, 2005). To calculate d' from pH and pF , $P_{c_{\max}}$ is first estimated using:

$$P_{c_{\max}} = \Phi\{[z(pH) - z(pF)]/2\} \quad (6.28)$$

and the result substituted into Equation 6.27. This calculation is performed by **PAL_SDT_1AFCsameDiff_IndMod_PHFtoDP**. The same routine also calculates the observer's criterion using $C = -0.5[z(pH) + z(pF)]$. The reverse calculation (pH and pF from d' and C) is performed by **PAL_SDT_1AFCsameDiff_IndMod_DPttoPHF**.

6.3.6 Calculation of d' for Match-to-Sample Tasks

6.3.6.1 Calculation of d' for 2AFC Match-to-Sample: Independent Observation Model

The computation of d' for the 2AFC match-to-sample task under the independent observation model parallels that of the 1AFC same-different task. According to Macmillan and Creelman (2005), who refer to the task as ABX, P_c for an unbiased observer is given by:

$$P_c = \Phi(d'/\sqrt{2}) \cdot \Phi(d'/2) + \Phi(-d'/\sqrt{2}) \cdot \Phi(-d'/2) \quad (6.29)$$

We refer readers to Macmillan and Creelman (2005) for the derivation of this equation. The calculation is performed by **PAL_SDT_2AFCmatchSample_IndMod_DPttoPC**. The inverse routine **PAL_SDT_2AFCmatchSample_IndMod_DPttoPC** employs an iterative search procedure using Equation 6.29 to obtain d' from P_c . If the raw data are hits and false alarms, defined according to the rule for a conventional 2AFC task, **PAL_SDT_2AFCmatchSample_IndMod_PHFtoDP** first calculates $P_{c_{\max}}$ and then d' by iterative search of Equation 6.29. **PAL_SDT_2AFCmatchSample_IndMod_DPttoPHF** performs the reverse calculations.

6.3.6.2 Calculation of d' for 2AFC Match-to-Sample: Differencing Model

For the 2AFC match-to-sample differencing model, the observer is assumed to encode the difference in sensory magnitude between the Sample and each of the Match stimuli, and choose the Match with the smallest absolute Sample-minus-Match difference. According to Macmillan and Creelman (2005) the differencing strategy, as with the same-different task, is the more likely to be adopted in a roving experiment where many different stimuli are presented during

a session. Macmillan and Creelman (2005) have derived the following equation for the unbiased observer:

$$P_c = \Phi(d'/\sqrt{2}) \cdot \Phi(d'/\sqrt{6}) + \Phi(-d'/\sqrt{2}) \cdot \Phi(-d'/\sqrt{6}) \quad (6.30)$$

Palamedes performs the calculation in

PAL_SDT_2AFCmatchSample_DiffMod_DPtoPC. The inverse routine **PAL_SDT_2AFCmatchSample_DiffMod_PCToDP** calculates d' by iterative search of Equation 6.30. If the data are hits and false alarms, d' and criterion C can be obtained using **PAL_SDT_2AFCmatchSample_DiffMod_PHFtoDP**, whose inverse is **PAL_SDT_2AFCmatchSample_DiffMod_DPToPHF**. The calculations in these routines parallel those for the 2AFC match-to-sample independent observation model.

6.3.6.3 Calculation of d' for M-AFC Match-to-Sample

For match-to-sample tasks in which $M > 2$, the Palamedes routines **PAL_SDT_MAFCmatchSample_DPtoPC** and **PAL_SDT_MAFCmatchSample_PCToDP** assume an unbiased observer and a differencing strategy. In the first routine, P_c is computed by Monte Carlo simulation rather than by equation. The simulation works as follows. Let S_1 and S_2 represent the two signal distributions separated by d' . Let S_1 be the Sample stimulus. Therefore, the M Match stimuli comprise one S_1 and $M - 1$ S_2 s. On each “trial” of the simulation, two random samples (don’t confuse a random “sample” with the “Sample” stimulus!) are selected from S_1 (one for the Sample stimulus and one for the Match stimulus), and $M - 1$ random samples are selected from S_2 (the other Match stimuli). The absolute difference between the Sample S_1 and each of the M Match stimuli is then calculated. If the absolute difference between the Sample S_1 and the Match S_1 is smaller than all of the absolute differences between the Sample S_1 and Match S_2 s, then by the differencing strategy the trial is scored “correct,” otherwise “incorrect.” The process is then repeated over a large number of trials and the overall proportion correct calculated across trials. Note that if the sample stimulus was designated to be S_2 the result would be expected to be the same, so the choice of S_1 or S_2 as the Sample is arbitrary. Because the routine uses Monte Carlo simulation, the computed P_c will not be identical each time, but should be accurate to about two decimal places. The routine is also relatively slow owing to the large number of trial simulations involved. The inverse routine that calculates d' from P_c is especially slow since it involves an iterative search of the forward routine.

6.3.7 Calculation of d' for M-AFC Oddity Tasks

The Palamedes routine **PAL_SDT_MAFCodidity_DPtoPC** also assumes an unbiased observer and a differencing strategy. It calculates P_c by Monte Carlo simulation following the method described by Craven (1992). Let S_1 and S_2 represent the

a session. Macmillan and Creelman (2005) have derived the following equation for the unbiased observer:

$$P_c = \Phi(d'/\sqrt{2}) \cdot \Phi(d'/\sqrt{6}) + \Phi(-d'/\sqrt{2}) \cdot \Phi(-d'/\sqrt{6}) \quad (6.30)$$

Palamedes performs the calculation in

PAL_SDT_2AFCmatchSample_DiffMod_DPtoPC. The inverse routine **PAL_SDT_2AFCmatchSample_DiffMod_PCToDP** calculates d' by iterative search of Equation 6.30. If the data are hits and false alarms, d' and criterion C can be obtained using **PAL_SDT_2AFCmatchSample_DiffMod_PHFtoDP**, whose inverse is **PAL_SDT_2AFCmatchSample_DiffMod_DPToPHF**. The calculations in these routines parallel those for the 2AFC match-to-sample independent observation model.

6.3.6.3 Calculation of d' for M-AFC Match-to-Sample

For match-to-sample tasks in which $M > 2$, the Palamedes routines **PAL_SDT_MAFCmatchSample_DPtoPC** and **PAL_SDT_MAFCmatchSample_PCToDP** assume an unbiased observer and a differencing strategy. In the first routine, P_c is computed by Monte Carlo simulation rather than by equation. The simulation works as follows. Let S_1 and S_2 represent the two signal distributions separated by d' . Let S_1 be the Sample stimulus. Therefore, the M Match stimuli comprise one S_1 and $M - 1$ S_2 s. On each “trial” of the simulation, two random samples (don’t confuse a random “sample” with the “Sample” stimulus!) are selected from S_1 (one for the Sample stimulus and one for the Match stimulus), and $M - 1$ random samples are selected from S_2 (the other Match stimuli). The absolute difference between the Sample S_1 and each of the M Match stimuli is then calculated. If the absolute difference between the Sample S_1 and the Match S_1 is smaller than all of the absolute differences between the Sample S_1 and Match S_2 s, then by the differencing strategy the trial is scored “correct,” otherwise “incorrect.” The process is then repeated over a large number of trials and the overall proportion correct calculated across trials. Note that if the sample stimulus was designated to be S_2 the result would be expected to be the same, so the choice of S_1 or S_2 as the Sample is arbitrary. Because the routine uses Monte Carlo simulation, the computed P_c will not be identical each time, but should be accurate to about two decimal places. The routine is also relatively slow owing to the large number of trial simulations involved. The inverse routine that calculates d' from P_c is especially slow since it involves an iterative search of the forward routine.

6.3.7 Calculation of d' for M-AFC Oddity Tasks

The Palamedes routine **PAL_SDT_MAFCodidity_DPtoPC** also assumes an unbiased observer and a differencing strategy. It calculates P_c by Monte Carlo simulation following the method described by Craven (1992). Let S_1 and S_2 represent the

two signal distributions separated by d' . Assume that S_1 is the oddity. Therefore, there are $M - 1$ non-oddity S_2 s. On each trial of the simulation, a random sample is selected from the S_1 oddity and each of the $M - 1$ non-oddity S_2 s. Consider that every random sample is a possible oddity. Recall that the decision rule is to select the alternative most different from the average of all the alternatives. We therefore calculate the absolute difference between each sample and the average of all the M samples. According to the decision rule, the sample selected to be the oddity is the one with the biggest absolute difference. If this sample is from S_1 then the trial is scored "correct," else "incorrect." The process is repeated over a large number of trials and the proportion correct calculated across trial. As with the routines for M -AFC match-to-sample, the computed P_c will not be identical each time, but should be accurate to about two decimal places. The routines are also relatively slow owing to the large number of trial simulations involved.

Further Reading

The best starting points for SDT are McNicol (2004), Chapters 5–8 of Gescheider (1997) and Macmillan & Creelman (2005). The most comprehensive treatment of SDT that is accessible to the non-expert, is Macmillan & Creelman (2005). More mathematical treatments can be found in Wickens (2002) and Green & Swets (1974). Further details of the computation of d' for the same-different tasks can be found in Macmillan et al. (1977).

Exercises

1. Consider the $M > 2$ versions of the standard forced-choice, oddity and match-to-sample tasks. The Palamedes routines for the M -AFC versions of these tasks assume that there are just two stimuli, S_1 and S_2 , and that the observer is unbiased and employs the following decision rules: for the standard forced-choice task select the alternative with the largest stimulus magnitude; for the oddity task select the alternative most different from the mean of all the alternatives; for the match-to-sample task select the match most similar to the sample. For a given d' , which task would you expect to produce the biggest and which the smallest P_c ? Write a script using the Palamedes routines to plot P_c against M for a given d' for each task to test your predictions.
2. The following exercise emerged from discussions with Mark Georgeson. Table 6.2 presents the results of an experiment aimed at measuring a psychometric function of proportion correct against stimulus magnitude using a standard 2AFC task. The experimenter is interested in the effects of bias on the estimates of the threshold and slope of the psychometric function, so the results are presented in terms of proportion hits pH and proportion false alarms pF , as calculated according to the rules in Section 6.2.4.

TABLE 6.2 Results of a hypothetical experiment aimed at deriving a psychometric function using a standard 2AFC task

Stimulus magnitude	<i>pH</i>	<i>pF</i>
1	0.61	0.53
2	0.69	0.42
3	0.79	0.33
4	0.88	0.18
5	0.97	0.06
6	0.99	0.03

Use the appropriate Palamedes routines to calculate d' , criterion C , and proportion correct P_c , for each pair of pH and pF . Then calculate the $P_{c_{\max}}$ for each stimulus magnitude that would be expected if the observer was unbiased (see Section 6.2.8). Plot psychometric functions of both P_c and $P_{c_{\max}}$ against stimulus magnitude (Chapter 4) and obtain estimates of the thresholds and slopes of the functions. Are the thresholds and slopes significantly different for the two functions (see Chapter 8)?

References

- Craven, B. J. (1992). A table of d' for M-alternative odd-man-out forced-choice procedures. *Perception & Psychophysics*, 51, 379–385.
- Elliot, P. B. (1964). Tables of d' . In J. A. Swets (Ed.), *Signal Detection and Recognition by Human Observers*. New York: Wiley.
- Gescheider, G. A. (1997). *Psychophysics: The Fundamentals*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Green, D. A., & Swets, J. A. (1974). *Signal Detection Theory and Psychophysics*. Huntington, New York: Krieger.
- Kingdom, F., Moulden, B., & Hall, R. (1987). Model for the detection of line signals in visual noise. *Journal of Optical Society of America A-Optics Image Science and Vision*, 4, 2342–2354.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Macmillan, N. A., Kaplan, H. L., & Creelman, C. D. (1977). The psychophysics of categorical perception. *Psychological Review*, 84, 452–471.
- McNicol, D. (2004). *A Primer of Signal Detection Theory*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Wickens, T. D. (2002). *Elementary Signal Detection Theory*. Oxford, New York: Oxford University Press.
- Yeshurun, Y., Carrasco, M., & Maloney, L. T. (2008). Bias and sensitivity in two-interval forced procedures. *Vision Research*, 48, 1837–1851.