Signal Detection Theory (SDT)

Hervé Abdi[1]

1 Overview

Signal Detection Theory (often abridged as SDT) is used to analyze data coming from experiments where the task is to categorize ambiguous stimuli which can be generated either by a known process (called the *signal*) or be obtained by chance (called the *noise* in the SDT framework). For example a radar operator must decide if what she sees on the radar screen indicates the presence of a plane (the signal) or the presence of parasites (the noise). This type of applications was the original framework of SDT (see the founding work of Green & Swets, 1966) But the notion of signal and noise can be somewhat metaphorical is some experimental contexts. For example, in a memory recognition experiment, participants have to decide if the stimulus they currently see was presented before. Here the signal corresponds to a familiarity feeling generated by a memorized stimulus whereas the noise corresponds to a familiarity feeling generated by a new stimulus.

The goal of detection theory is to estimate two main parameters from the experimental data. The first parameter, called d', indicates the strength of the signal (relative to the noise). The second parameter called C (a variant of it is called β), reflects the strategy

[1]In: Neil Salkind (Ed.) (2007). *Encyclopedia of Measurement and Statistics.* Thousand Oaks (CA): Sage.
Address correspondence to: Hervé Abdi
Program in Cognition and Neurosciences, MS: Gr.4.1,
The University of Texas at Dallas,
Richardson, TX 75083–0688, USA
E-mail: herve@utdallas.edu http://www.utd.edu/~herve

Table 1: The four possible types of response in SDT

	DECISION: (PARTICIPANT'S RESPONSE)	
REALITY	Yes	No
Signal Present	Hit	Miss
Signal Absent	False Alarm (FA)	Correct Rejection

of response of the participant (*e.g.,* saying easily yes rather than no). SDT is used in very different domains from psychology (psychophysics, perception, memory), medical diagnostics (do the symptoms match a known diagnostic or can they be dismissed are irrelevant), to statistical decision (do the data indicate that the experiment has an effect or not).

2 The Model

It is easier to introduce the model with an example, so suppose that we have designed a face memory experiment. In the first part of the experiment, a participant was asked to memorize a list of faces. At test, the participant is presented with a set of faces one at a time. Some faces in the test were seen before (these are *old* faces) and some were not seen before (these are *new* faces). The task is to decide for each face if this face was seen (response *Yes*) or not (response *No*) in the first part of the experiment.

What are the different types of responses? A Yes response given to an old stimulus is a correct response, it is called a *Hit*; but a Yes response to a new stimulus is a mistake, it is called a *False Alarm* (abbreviated as FA). A No response given to a new stimulus is a correct response, it is called a *Correct Rejection*; but a No response to an old stimulus is a mistake, it is called a *Miss* (abbreviated as FA). These four types of response (and their frequency) can be organized as shown in Table 1.

The relative frequency of these four types of response are not all independent. For example when the signal is present (first row of Table 1) the proportion of Hits and the proportion of Misses add up to one (because when the signal is present the subject can say either Yes or No). Likewise when the signal is absent, he proportion of FA and the proportion of Correct Rejection add up to one. Therefore all the information in a Table such as Table 1) is given by the proportion of Hits and FAs.

Even though the proportions of Hits and FAs provide all the information in the data, these values are hard to interpret because they crucially depend upon two parameters. The *first* parameter is the *difficulty* of the task: The easier the task the larger the proportion of Hits and the smaller the proportion of FAs. When the task is easy, we say that the signal and the noise are well separated, or that there is a large distance between the signal and the noise (conversely, for a hard task, the signal and the noise are close and the distance between them is small). The *second* parameter is the *strategy* of the participant: A participant who always says No will never commit a FA; on the other hand, a participant who always says Yes is guaranteed all Hits. A participant who tends to give the response Yes is called *liberal* and a participant who tends to give the response No is called *conservative*.

3 The SDT model

So, the proportions of Hits and FAs reflect the effect of two underlying parameters: the first one reflects the separation between the signal and the noise and the second one the strategy of the participant. The goal of SDT is to *estimate* the value of these two parameters from the experimental data. In order to do so, SDT creates a *model* of the participant's response. Basically the SDT model assumes that the participant's response depends upon the intensity of a hidden variable (*e.g.,* familiarity of a face) and that the participant responds Yes when the value of this variable for the stimulus is larger than a predefined threshold.

SDT also assumes that the stimuli generated by the noise condition vary naturally for that hidden variable. As is often the case elsewhere, SDT, in addition, assumes that the hidden variable values for the noise follow a normal distribution. Recall at this point, that when a variable x follows a Gaussian (a.k.a Normal) distribution, this distribution depends upon two parameters: the mean (denoted μ) and the variance (denoted σ^2). It is defined as:

$$\mathscr{G}(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} . \qquad (1)$$

In general within the SDT framework the values of μ and σ are arbitrary and therefore we choose the simpler values of $\mu = 0$ and $\sigma = 1$ (other values will give the same results but with more cumbersome procedures). In this case, Equation 1 reduces to

$$\mathscr{N}(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\} . \qquad (2)$$

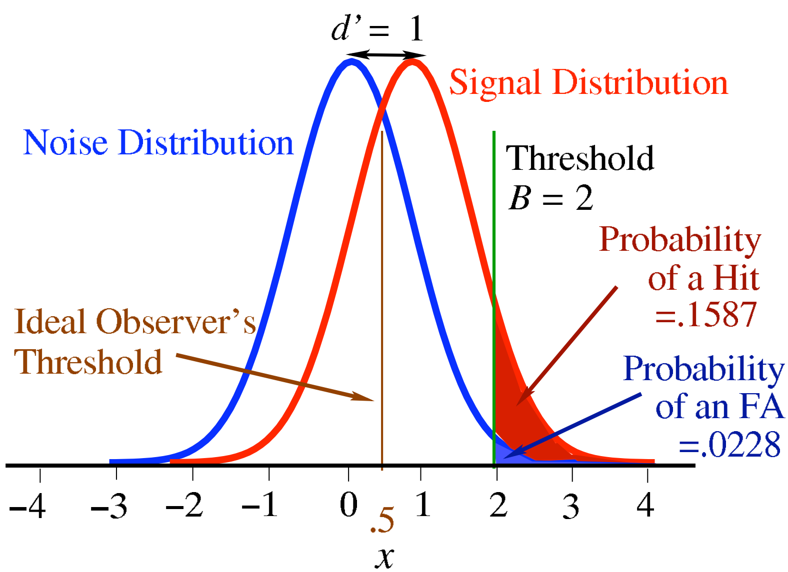

Figure 1: The model of SDT.

Finally, SDT assumes that the signal is *added* to the noise. In other words, the distribution of the values generated by the signal condition has the *same* shape (and therefore the same variance) as the noise distribution.

Figure 1 illustrates the SDT model. The x-axis shows the intensity of underlying hidden variable (*e.g.,* familiarity for the face example). As indicated above, the distribution of the noise is centered at zero (*i.e.,* mean of the noise is equal to zero, with a standard deviation of 1. So, the standard deviation of the noise is equivalent to the unit of measurement of x. The distribution of the signal is identical to the noise distribution, but it is moved to the right of the noise distribution. The distance between the signal and the noise distributions corresponds to the effect of the signal (this is the quantity that is added to the noise distribution in order to get the signal distribution): this distance is called d'. Because the mean of the noise distribution is zero, d' is equal to the mean of the signal distribution.

The strategy of the participant is expressed via the choice of the threshold. There are several ways of expressing the position of this threshold, among the possible candidates we will mention four of them denoted B, D, C and β. The first quantity B (sometimes called ϑ) gives the position of the threshold on the x-axis. In the example illustrated in Figure 1, this value is equal to 2, and so the participant corresponding to this figure has decided that any stimulus with a value of x larger than 2 comes from the signal distribution and is given the response Yes. The position of the threshold can also be given relative to the signal distribution (because the noise has zero mean, B is the distance of the threshold relative to the noise distribution), as the mean of the signal is equal to d' we can compute D as $D = d' - B$ (a value equal to 1 in our example).

The most popular way of expressing the location of the threshold, however, is neither from the distribution of the noise nor the distribution of the signal but relative to what is called the *ideal observer.* The ideal observer minimizes conjointly the probability of a Miss and of an FA. When each type of errors has the same cost, the criterion of the ideal observer is positioned on the average of the means of the signal and the noise distribution. In our example, the threshold of the ideal observer would be equal to $\frac{1}{2}d' = \frac{1}{2} = .5$. The value of C is the distance from the actual threshold to the ideal observer, it can computed as $C = B - \frac{d'}{2} = 2 - .5 = 1.5$. The sign of C reveals the participant's strategy: when $C = 0$, we have the ideal observer; when C is negative the participant is *libéral* (*i.e.,* responds Yes more often than the ideal observer); when C is positive the participant is *conservative* (*i.e.,* responds No more often than the ideal observer).

An alternative way of expressing the position of the participant's criterion is given by the quantity called β. It corresponds to the ratio of the height of the signal distribution to the noise distribution for the value of the threshold. Because the distributions of the noise and the signal are normal with variance equal to one, we can compute β from Equation 2 as:

$$\beta = \frac{\mathcal{N}(D)}{\mathcal{N}(B)} = \frac{\mathcal{N}(1)}{\mathcal{N}(2)} = \frac{.2420}{.0540} \approx 4.4817 . \tag{3}$$

Some rewriting can show that Equation 3 can be rewritten as

$$\beta = \exp\left\{d' \times C\right\} . \tag{4}$$

The quantity β has the advantage of being a *likelihood ratio* and can be used to interpret SDT within a statistical framework. For practical reasons, it is often easier to compute the logarithm of β, for example from Equation 4, we get

$$\ln \beta = d' \times C = 1 \times 1.5 = 1.5 . \tag{5}$$

The model illustrated by Figure 1 generates a specific pattern of response probabilities which can be computed from integrating the normal distribution. So, for example, the probability of a FA is obtained as the probability (*i.e.,* area under the normal distribution) of finding a value larger than 2 with a normal distribution of mean 0 and variance 1 (this can be computed with most statistical packages or from Tables such as the ones given in Abdi, 1987). This quantity is also called the probability *associated* to the value 2, in our example it is equal to .0228. Along the same lines, the probability of a Hit is obtained as the probability (*i.e.,* area under the normal distribution) of finding a value larger than 2 with a normal distribution of mean 1 (*i.e.,* the mean of the signal) and variance

Table 2: The probability or the four possible types of response according to to Figure 1.

	DECISION: (PARTICIPANT'S RESPONSE)		
REALITY	Yes	No	Total
Signal Present	Hit	Miss	
	Pr{Hit}=.1587	Pr{Miss}=.8413	1
Signal Absent	False Alarm (FA)	Correct Rejection	
	Pr{FA}=.0228	Pr{Correct Rejection}=.9772	1

1, this is equivalent of finding the probability (*i.e.,* area under the normal distribution) of finding a value larger than $2 - 1 = 1$ with a normal distribution of mean $1 - 1 = 0$ and variance 1. This value is equal to .1587.

4 SDT in practice

The previous example was describing the performance of a participant who behaved according to the SDT model. However, in practice we do not known the values of the parameters of SDT, but we want to *estimate* them from the performance of the participants. In an experimental paradigm the only observable quantities are the participant's responses from which we can derive the number of hits and FA's.

To illustrate this problem suppose that we want to evaluate the performance of a wine taster whose task is to detect if a wine labelled as made from "Pinot Noir" has been tempered by the addition of some Gamay (generally considered an inferior grape). Here, the signal corresponds to presence of Gamay. Our wine taster tasted (blindfolded) twenty glasses of Pinot, (half of them tempered with

Table 3: The performance of a wine taster trying to identify Gamay in a Pinot Noir wine.

	DECISION: (TASTER'S RESPONSE)		
REALITY	Yes (Gamay)	No (Pure Pinot)	$\sum$
Signal Present	Hit	Miss	
(Gamay)	#{Hit}=9	#{Miss}=1	10
	Pr{Hit}=.9	#{Miss}=.1	1
Signal Absent	False Alarm (FA)	Correct Rejection	
(Pure Pinot)	#{FA}=2	#{Correct Rejection}=8	10
	Pr{FA}=.2	Pr{Correct Rejection}=.8	1

some Gamay and half without). The results are reported in Table 3, and show that the proportion of Hits and FAs are respectively .9 and .2. In order to find the values of d' and the criterion, we need to inverse the formulas given above (*i.e.,* Equation 3–5). We need one new notation: for a normal distribution with zero mean, we denote by Z_P the value of the normal distribution whose associated probability is equal to P (*e.g.,* $Z_{.025} = 1.96$). We denote Z_H et Z_{FA} the values corresponding to the proportions of Hits and FAs. With these new notations and after some (minor) algebraic manipulations we find the following set of formulas. The estimation of d' is obtained as

$$d' = Z_H - Z_{FA} = Z_{.9} - Z_{.2} = 1.28 - (-.84) = 2.12\,. \tag{6}$$

The estimation of C is obtained as

$$C = -\frac{1}{2}\left[Z_H + Z_{FA}\right] = -\left[Z_{.9} + Z_{.2}\right] = -\frac{1}{2}[1.28 - .84] = -.22\,, \tag{7}$$

and $\ln\beta$ is obtained as

$$\ln\beta = d' \times C = 2.12 \times -.22 = -.47 \tag{8}$$

(β is obtained as $\exp\{\ln \beta\} = .63$).

How to interpret these results? The taster is clearly (but not perfectly) discriminating between Pinots and tempered Pinots (as indicated by a d' of 2.12), this taster is also liberal (in case of doubt the taster will rather say that the wine has been tempered rather than not).

5 Bibliography

The classic work on SDT is Green and Swets (1966), a basic introduction is McNicol, D. (1972), two recent comprehensive references are Macmillan and Creelman (2005) and Wickens (2002).

References

[1] Abdi, H. (1987). *Introduction au traitement des données expérimentales*. Grenoble: Presses Universitaires de Grenoble.

[2] Green D.M., Swets, J.A. (1966). *Signal detection theory and psychophysics*. New York Wiley.

[3] Macmillan, N.A., Creelman, C.D. (2005). *Detection theory: A user's guide* (2nd edition). Mahwah (NJ): Erlbaum.

[4] McNicol, D. (1972). *A primer of signal detection theory*. London: George Allen & Unwin.

[5] Wickens, T.D. (2002). *Elementary signal detection theory*. Oxford: Oxford University Press.