Signal Detection Theory

From "Encyclopedia of Cognitive Science"

Signal detection theory is one of the first attempts by psychologists to model the processes involved in elementary perceptual recognition tasks. The theory's main feature is a rigorous distinction between performance limits due to the quality of information provided by the senses (sensitivity), and limits arising from decision-making strategies (bias).

INTRODUCTION

Imagine the following situation: a pathologist has been given a tissue sample from a male child to check for the presence of cancer. If she incorrectly concludes that the sample is cancerous, the child will unnecessarily undergo a battery of costly and potentially risky interventions. On the other hand, if she incorrectly concludes that there is no cancer present, the disease may be allowed to develop beyond the ability of medical science to treat it. Because this disease is relatively rare in children, the pathologist knows that the probability that the cells are cancerous is small. Before looking at the tissue, therefore, she might be fairly confident that it is noncancerous. Unfortunately, the signs of cancer in a tissue sample are not clear-cut. Every sample is unique; sometimes there are obvious, strong indications that cancer is present, at other times the signs are moderate or weak. How does the pathologist combine the complex visual information from examining the tissue sample with her prior knowledge about the likelihood of a cancer being present? How do we evaluate the performance of these expert decision-makers to determine whether they are making the best possible use of their knowledge about the probability of the disease in this situation and the visual perceptual information they glean from examining the sample?

SIGNAL DETECTION THEORY

The pathologist's dilemma is one of many examples of what psychologists call the 'signal detection problem'. Are the cells in the tissue sample cancerous (a signal) or is it a false alarm (noise)? There are two possible events (signal and noise) and two possible decisions (detect and reject).

During the Second World War, engineers working on radar detection systems developed a model to describe the processes involved in this particular type of decision-making problem. They used this model to study the behavior of radar operators who were responsible for distinguishing random blips on the radar screen from blips caused by an actual ship or aircraft (the signal detection problem). This mathematical model was soon noticed by psychologists and developed into a set of more detailed models and procedures, which have come to be referred to collectively as 'signal detection theory' (SDT).

According to this general theory, there are two distinct stages in the detection process, assimilation of information from the environment (encoding) and selection of an appropriate response (decision-making). Detection errors occur either because the output of the encoding process is misleading (due to noise in the outside world or in the brain) or because the detector's decision-making strategy is flawed.

Although the basic principles of SDT are general enough to encompass more complex decision-making tasks, most applications have focused on the original problem, in which there are only two possible stimuli and two possible responses. In such situations (often called discrimination tasks), the four possible outcomes of a given trial during the experiment are 'hits' (correctly detect a signal), 'misses' (fail to detect a signal), 'correct rejections' (reject a non-signal event) and 'false alarms' (incorrectly identify a non-signal event as a signal event).

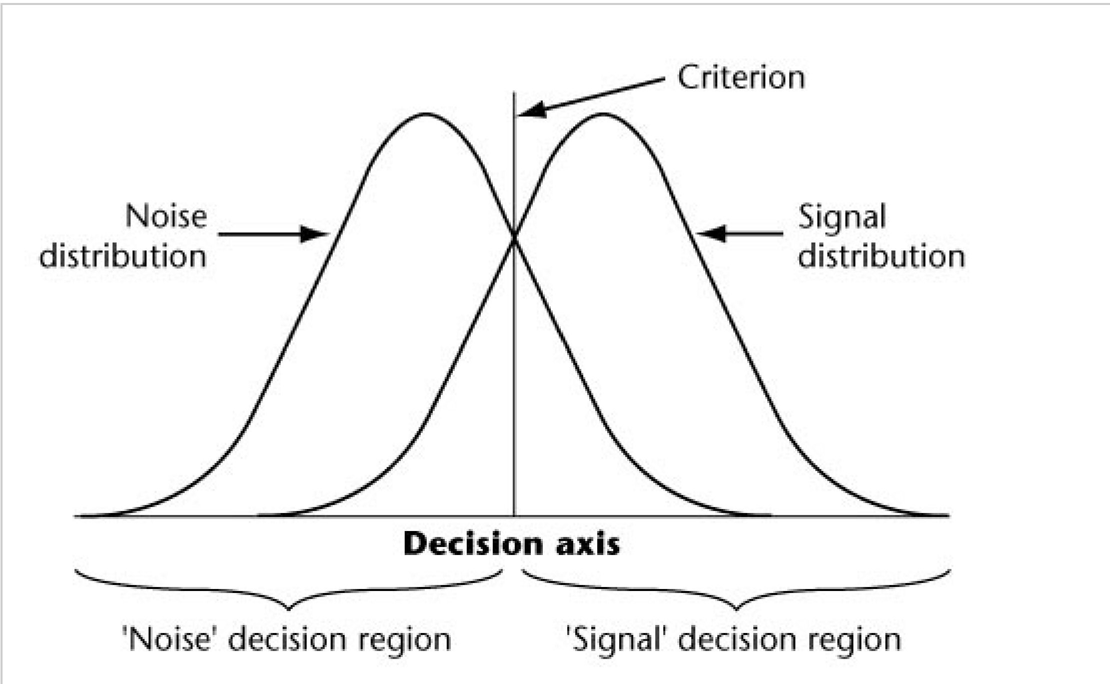
Since there are only two possible decisions, detect or reject, the proportion of hits on signal trials plus the proportion of misses on signal trials must equal 1. Similarly, the proportion of false alarms on noise trials plus the proportion of correct rejections on noise trials must equal 1. The performance of a subject can therefore be reduced to two values, one for each stimulus condition, and it is customary to report just the hit and false alarm rates.

SENSITIVITY AND BIAS

One of the basic assumptions of SDT is that all of the sensory information that could conceivably be used to come to a decision on a given trial can be represented as a single point on a one-dimensional continuum called the 'decision axis'. Points near the left extreme of the axis correspond to sensory information that constitutes compelling evidence that only background noise was present; and points near the right end are compelling evidence for a signal. Somewhere between these two extremes lies the point at which the sensory information is entirely inconclusive (the two stimuli or conditions are equally likely). In the pathologist's situation, the axis might be the number of what appear to her to be questionable cells in the tissue sample: many questionable cells might constitute very strong evidence that the sample is cancerous while very few questionable cells might indicate a noncancerous sample. The decision-maker must combine this sensory information with other information, such as the history of the patient.

Because of the random variability that is almost sure to exist either in the environment or in the brain (or both), the sensory effect of a stimulus should be expected to vary from trial to trial even if the stimulus stays constant (in the pathology example, different cancerous tissue samples would look different). However, some effects should be more common than others when they come from a single type of source. For technical reasons, the most popular version of SDT assumes that these relative frequencies for a given stimulus follow a standard probability distribution, the 'normal' or 'bell-shaped' curve, as illustrated in Figure 1.

Notice that each stimulus has its own distribution, because each stimulus should have its own set of effects that it will produce often (effects near the center, or mean, of the distribution) and its own set of effects that are possible but rare (effects in the tails of the distribution). The degree to which the distributions for two different stimuli overlap is a measure of the 'sensitivity' of the senses to the physical differences between the two stimuli. In general, sensitivity should decrease with increasing physical similarity of the stimuli being discriminated.



The distributions of effects on signal and noise trials in SDT. Notice that the two distributions have the same variance but different means. According to SDT, the observer changes his or her response strategy by shifting the criterion along the decision axis, thereby changing the boundary between the 'signal' and 'noise' response regions. In this example, the criterion is placed at the point of intersection between the two distributions and is therefore unbiased. Response biases affect the placement of the criterion (shifting it to the left or right), but not the distributions themselves. (Figure 1.)

Decision-making Strategies

The fact that the two distributions in Figure 1 overlap is a fundamental problem for the decision-maker: for any given sensory effect, the decision-maker cannot be sure which stimulus was presented. From the SDT point of view, this is the main reason why subjects make errors in discrimination tasks. On the other hand, the sensory effect does provide some information: some effects are more likely to have been caused by the signal stimulus and others are more likely to have been caused by the noise stimulus. The decision-maker's task is therefore to find a 'decision rule' that maps each effect to its most probable cause. SDT assumes that, to accomplish this, a 'criterion' is set somewhere on the decision axis.

Intuitively, the best place to put this criterion might seem to be the point at which the two distributions intersect; i.e. where the relative frequency of the sensory effect is identical on signal and noise trials. This is valid if the signal and noise trials occur with equal frequency during the experiment (i.e., the 'base rates' of the stimuli are equal). However, if the base rates are unequal, the optimal placement (maximizing the percentage of correct responses) will be somewhere to the left or right of the point of intersection of the two distributions, depending on which stimulus is presented more often.

To see why the criterion should shift in accordance with the base rates, consider the case in which the signal stimulus will never be presented (when its base rate is zero): the optimal decision-maker would always respond 'noise', no matter what sensory effect is present. In terms of the model, the criterion in this case would be shifted to the extreme right of the decision axis so that all of the sensory effects are mapped to the noise response.

For each pair of base rates, there is one and only one position on the decision axis that would lead to optimal performance: all of the others are suboptimal to some degree. Any criterion placed at any point other than the point of intersection (which may or may not be suboptimal, depending on the base rates) is called a 'biased' decision rule.

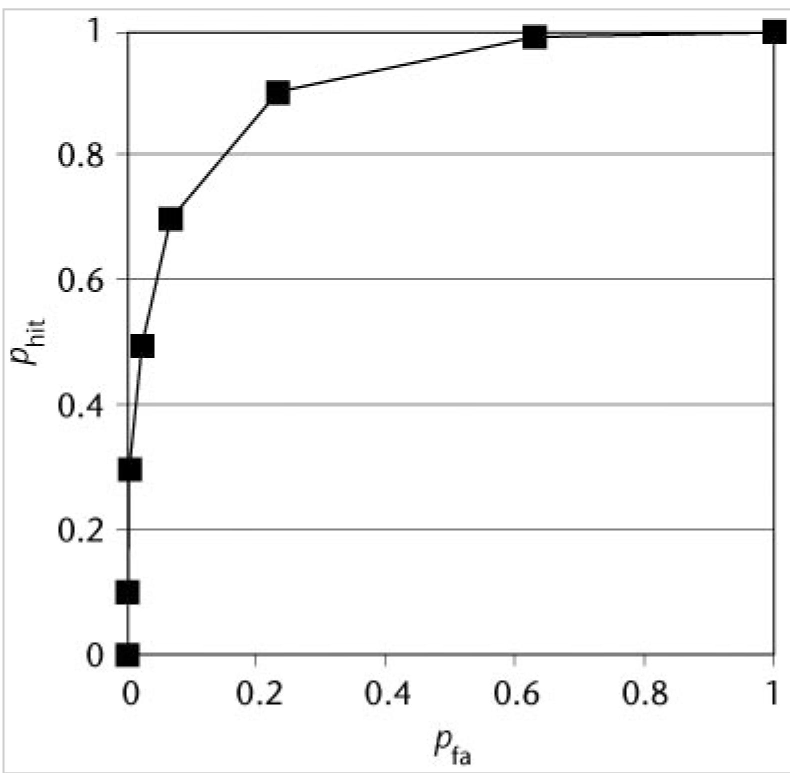
CALCULATING SENSITIVITY

For any set of base rates, the SDT model illustrated in Figure 1 can be fitted to the data (i.e. to the hit and false alarm rates) to estimate the two parameters of the model, which are the distance d' between the means of the two distributions (in standard deviation units) and the position X_C of the criterion. Assuming a standard deviation of 1, the formulas for these two parameters are: (1)$d' = \Phi^{-1}(p_{\text{hit}}) - \Phi^{-1}(p_{\text{fa}})$ (2)$X_C = -\Phi^{-1}(p_{\text{fa}})$ where p_{hit} and p_{fa} are the hit and false alarm rates, respectively, and Φ^{-1} is the inverse cumulative distribution function for the standard normal distribution. The values of these inverse functions can be obtained from tables or from statistical software packages.

ROC CURVES

In principle, there is no reason why the two distributions of effects should always be normal, or should differ only in their mean values. In fact, empirical studies often suggest that the signal distribution is more variable than the noise distribution. Partly for these reasons, detection theorists often estimate sensitivity using a different approach that incorporates the same basic principles of the detection model but does not assume anything about the particular shapes of the distributions. To do this, it is necessary to estimate the hit and false alarm rates for several different positions of the decision criterion. These pairs of estimates are then presented in a scatter diagram, with the false alarm rate on the abscissa and the hit rate on the ordinate. An example of such a 'receiver operating characteristic' (ROC) curve is shown in Figure 2. In addition to illustrating how different degrees of bias towards one of the two responses would affect a decision-maker's performance, such plots can also provide a convenient estimate of sensitivity, because the area underneath a plot will generally increase as the sensitivity of the subject increases.

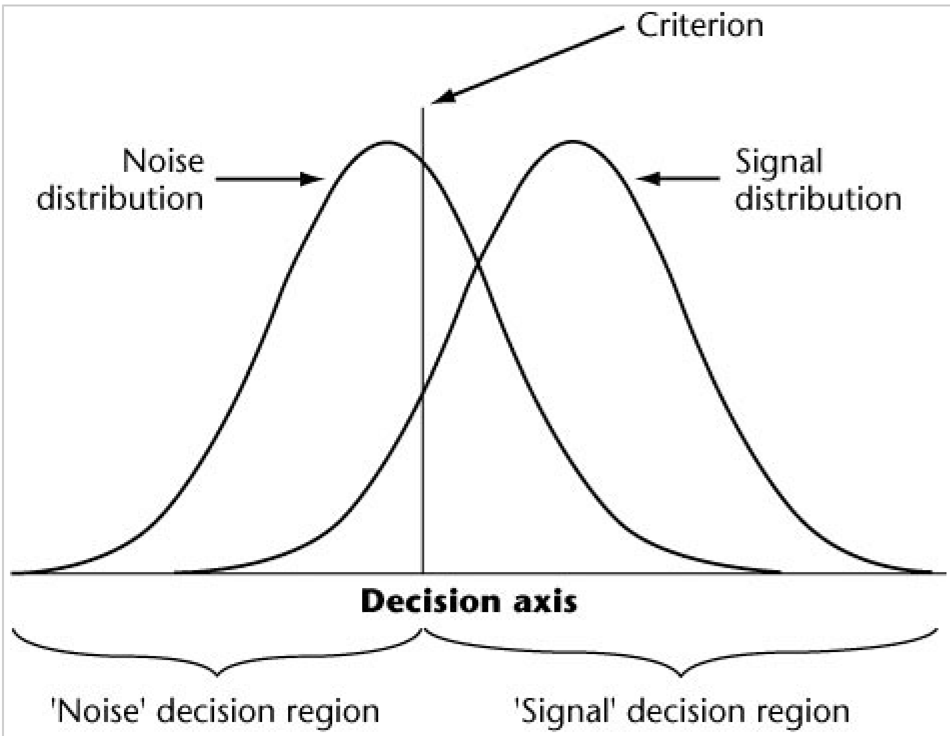
Of course, in order to estimate the ROC curves at more than one point, the investigator needs more estimates of the hit and false alarm rates than are available from a standard discrimination task. One approach is to run several experiments using different base rates, presumably inducing the subjects to shift their decision criteria to the left or right depending on the relative frequency of the signal. However, since this approach is more costly in both time and effort, most investigators prefer an alternative method, which is to run a single condition with equal base rates but to ask the subjects to report both the stimulus that they think was presented and their degree of confidence in this. In effect, this is equivalent to asking the subjects to report the position of the sensory effect on the decision axis as well as the response they wish to assign to this effect. From this extra information it is possible to estimate the ROC curve.



ROC curve from a hypothetical discrimination experiment. Points along the curve correspond to conditions with different signal base rates. (Figure 2.)

APPLICATIONS OF SDT

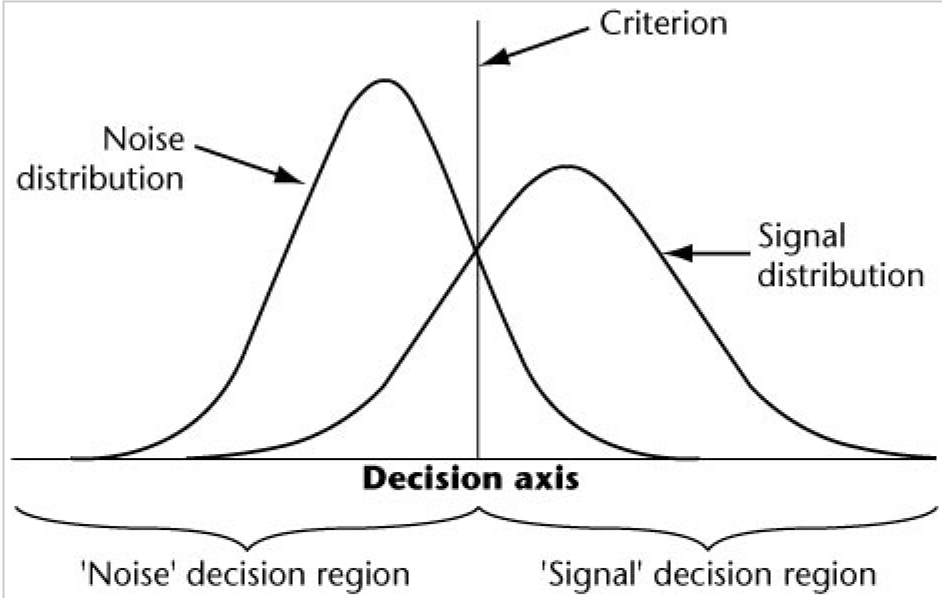
SDT has been adopted as a framework for studying human performance in many different areas of basic and applied research, including perception and memory, decision-making, radar monitoring, airport security, and personnel evaluation. Because the need to discriminate between two possible states of the physical world from imperfect (but not useless) information arises in so many situations in the laboratory and in the workplace, it is important to have a rigorous means of describing these kinds of activities and the effects that decision-making biases can have on them. When its underlying assumptions are known to be satisfied, SDT can be a useful tool to help assess decision-making styles and distinguish these styles from other aspects of the skills involved in a discrimination task. In these cases, it provides a legitimate method of identifying and helping to correct potentially harmful biases of decision-makers in critical decision situations. In addition, SDT is often used as a basis for testing theories that make predictions about factors that either should or should not affect the inherent difficulty of a task (rather than merely the strategies that subjects will choose to employ), and to summarize in an efficient way the effects of these factors on performance.



An SDT model with a biased decision rule. The criterion has been shifted to the left of the point of intersection between the two distributions, indicating a bias towards the signal response (compared to the unbiased rule, more of the effects on the decision axis are mapped to the signal response). (Figure 3.)

However, these applications of the model have little or no scientific merit if the model's representation of behavior is inaccurate. It is important, therefore, to test the model's assumptions in a rigorous way before accepting them. Some of the most convincing evidence in favor of the model comes from manipulations that according to SDT should be directly related to the decision-making process (the placement of the criterion) and unrelated to the sensory encoding process (and hence the sensitivity level). In particular, if the noise and signal trials are not presented with equal frequency (the base rates are unequal), or if a 'miss' is, say, more costly than a 'false alarm', the optimal decision-maker will shift the criterion away from the point of intersection. If the criterion is shifted to the right, the hit and false alarm rate should both decrease; and many studies have shown that this is precisely what occurs when the frequency of the noise stimulus (or the penalty for a false alarm) is increased. In other words, this fundamental prediction of the SDT model is confirmed by empirical data.

SDT also makes some other predictions, however, that have been shown to be violated. If the criterion is biased towards, say, the signal response, as in the example shown in Figure 3, the theory predicts that the relative frequency of low-confidence signal responses (effects immediately to the right of the criterion) should be lower on signal trials than on noise trials. This prediction is consistently violated in discrimination experiments. These two relative frequencies convergence to the same value as the subjects' confidence decreases (the likelihood ratio at the criterion is 1).



A modification of the SDT model that is consistent with empirical data. Instead of the decision criterion shifting, the relative variances of the two distributions change as the bias increases. Higher base rates of a stimulus lead to a distribution with smaller variance. In this example, increasing the base rate of the noise stimulus decreased the variance of the noise distribution and increased the variance of the signal distribution. The decision rule is unbiased, but the change in variance accounts for the relationship between hit and false alarm rates when base rates are manipulated. (Figure 4.)

In order to explain the covariance of the hit and false alarm rates under base rate manipulations, while also explaining this direct empirical evidence that the decision rule is always unbiased, the SDT model needs to be modified so that the shapes of the distributions change as the base rates are manipulated. An example is shown in Figure 4. Instead of the criterion shifting, the relative variances of the two distributions change, with a higher base rate for a given stimulus leading to lower variance in its associated distribution of sensory effects. Obviously, this is a very different kind of bias, and one which eliminates the 'shifting criterion' representation of decision-making processes in SDT. It is not yet clear exactly what causes the distributions to change shape in the manner shown in Figure 4, or what consequences this effect has for measurement of discrimination skills. More research is needed to answer this important question. Meanwhile, applications of the traditional SDT models should be regarded sceptically.

Further Reading

- Green, D M and Swets, J A (1966) Signal Detection Theory and Psychophysics. New York, NY: John Wiley.
- Balakrishnan, J D and MacDonald, J A (2001) Alternatives to signal detection theory. In: Karwowski, W (ed.) International Encyclopedia of Ergonomics and Human Factors. London, UK: Taylor & Francis.
- Swets, J A (1998) Enhancing diagnostic decisions. In: Hoffman, R R, Sherrick, M F et al. (eds) Viewing Psychology as a Whole: The Integrative Science of William N. Dember. Washington, DC: American Psychological Association.

JUSTIN A MACDONALD
PURDUE UNIVERSITY, WEST LAFAYETTE, , INDIANA, USA
J D BALAKRISHNAN
PURDUE UNIVERSITY, WEST LAFAYETTE, , INDIANA, USA