[redacted]

R67-86

Acta Psychologica **27** (1967) 390–399; © *North-Holland Publishing Co., Amsterdam*

DETECTABILITY THEORY AND THE INTERPRETATION OF VIGILANCE DATA

M. M. TAYLOR

Defence Research Medical Laboratories, Toronto, Ontario, Canada [1,2]

ABSTRACT

The concept of subjective probability forms the basis of a brief summary of the theory of discrimination. The concept of likelihood ratio is considered as an interpretive convenience, rather than as a conceptual necessity, and is used to introduce the familiar ROC curve. Interpretation of vigilance data is discussed in terms of the expected form of the relevant ROC curves. For the detection of signals in a steady background, the typical paradigm of a vigilance experiment, the ROC curve may be severely skewed. When the data depend on a single operating point, the tabulated indices of detectability, d', and of caution, β, may give misleading impressions of actual detection behavior. The tabulated value of β will always be higher than the true β, usually drastically so, while the tabulated value of d' will usually, but not always, be higher than the true value of d'. The tabulated values will show correlation, which may be either positive or negative, across observers in situations where the true values of d' and β would show no correlation. Though the tabulated measures may be of dubious value, the concepts of detection theory remain useful in the analysis of vigilance.

1. Introduction

There is disagreement about how many detection theory parameters may be usefully applied to vigilance data. Broadbent and Gregory (1963, 1965) used both d', the index of detectability, and β, the index of caution. Their later paper defended the use of both indices. Mackworth and Taylor (1963) used d', but declined to use β on the grounds that it seemed more sensitive to the inevitable violations of detection theory assumptions. Wiener et al. (1964) used neither, feeling (Wiener, E.L. personal communication, 1964) that the assumptions were sufficiently badly violated, and the

PD-3063

[1] During the academic year 1966–7 at the Department of Communication, University of Keele, Staffs, England.

[2] This is Defence Research Medical Laboratories Research Paper No. 655.

data sufficiently disturbing in detail, that neither index could be used. This paper attempts a partial resolution of the problems arising from use of detectability theory in the context of vigilance.

2. Discrimination and subjective probability

Consider first the basic paradigm of detection and discrimination experiments. Two sorts of events are possible. Call them oranges and lemons. During an observation interval, either an orange or a lemon is presented. The observer does not know exactly which, since he only sees a distorted or noisy version of the event, but he must label a particular event as either an orange or a lemon.

The observer goes through some unspecified processes, winding up at some stage with an estimate of the probability that he observed an orange. This subjective probability is based partly on the *a priori* probability that the event would be an orange, and partly on the evidence provided by the observation. The evidence provided is different for each observation so that the probability arrived at by the observer is different each time an orange is actually presented. In principle, one can describe two separate subjective probability distributions, one for the cases in which an orange was presented, and another for the cases in which a lemon was presented.

How the observer responds once he has assessed the probability that an observed event was an orange depends on circumstances irrelevant to the actual discrimination process. If it is very important for the observer not to call any oranges lemons, then he should say 'orange' whenever he has the least suspicion that it might have been an orange. On the other hand, if it is equally important to be correct in either direction, then he should say 'orange' only if he thinks he has at least a 50–50 chance of being correct. For the same information gained from an observation, some situations will call for the response 'orange', others for 'lemon'. The choice belongs to the observer.

Assume the observer at least to order observations, so that if one observation leads him to respond 'orange' then another observation which yields a higher subjective probability in favour of orange will evoke the same response. This implies a fixed criterial subjective probability. Whenever the assessed probability of orange is greater than this criterion, then the subject should say 'orange'. Otherwise, he should say 'lemon'.

An observer can base his response only on his own assessment of the probability that the event was an orange. The experimenter is not able directly to measure this assessment by the observer. Besides, experimenters are often more interested in what information the observation gives the observer than in the final assessment. For both these reasons, conventional usage in detectability theory abandons subjective probability in favour of a more 'objective' measure, which is closely related to the value of the observation as evidence.

3. Likelihood ratios

It is convenient and elegant to consider the two event classes, oranges and lemons, together and symmetrically. The analyses use the ratio between the probability of oranges and the probability of lemons. Before the observation, this is the prior probability ratio; afterwards, it becomes the posterior probability ratio. One frequently invoked corrollary of Bayes' theorem (see, e.g., Edwards, Lindemann and Savage, 1963) is that posterior probability ratio is equal to prior probability ratio times a quantity known as the likelihood ratio.

The likelihood ratio is defined as the ratio between the probability of the observation given that the first alternative (orange) is true, and the probability of the observation given that the second alternative (lemon) is true. In theoretical analyses of decision behaviour, the likelihood ratio is normally obtained in the manner suggested by its definition, but in experiments, it is usually obtained indirectly from the relation of the prior and posterior probability ratios. In either case, the likelihood ratio provides a measure of the evidence given by an observation.

As long as the prior probability ratio is constant, the likelihood ratio distributions both reproduce the subjective posterior probability ratio distributions, with the abscissa divided by the prior probability ratio. Hence when the observer decides on a subjective probability ratio criterion, he is also selecting a likelihood ratio criterion.

The probability that the observer says 'orange' given that the event was an orange is determined by the proportion of the orange-contingent likelihood-ratio probability-density function above the criterion, while the probability that he says 'orange' when the event was a lemon is similarly determined by the proportion of the lemon-contingent function above the criterion. These represent all the observations from which the observer got more than enough evidence of orangeness to decide on 'orange' as his response. To determine exactly how much evidence the observer needs, it is necessary to determine the relative probabilities of obtaining exactly the criterion likelihood ratio when orange is true and when lemon is true. The criterion likelihood ratio, in other words, is the ratio between the heights of the two density functions at criterion.

The data do not give these heights; they give the integrated density above criterion. To determine the heights of the density functions, one must differentiate, which, in concept, implies the performance of two experiments identical except that the criteria differ infinitesimally. The integral of the orange density function and the integral of the lemon density function are both functions of the criterion placement. Since the ratio of the derivatives of two functions of the same variable with respect to that variable is equal to the derivative of one function with respect to the other—

$$(\mathrm{d}Y/\mathrm{d}X)/(\mathrm{d}Z/\mathrm{d}X) = \mathrm{d}Y/\mathrm{d}Z$$

—one can eliminate direct reference to the criterion placement, which is not given by data. The criterion likelihood ratio is the derivative of the probability of an 'orange' response to an orange event, with respect to the probability of an 'orange' response to a lemon event.

The determination of the likelihood ratio as the derivative of a functional relationship between two experimentally determined probabilities suggests that the function relating them is itself interesting. The entire function is described by permitting the criterion likelihood ratio to range from zero to infinity. When the criterion likelihood ratio is zero, each distribution lies entirely above it, so that both the probabilty of an 'orange' response to an orange event and the probability of an 'orange' response to a lemon event are unity. Similarly, when the criterion is infinity, both probabilities are zero. The function, known as the ROC curve, traces some sort of curve within the unit square, starting at (0,0) and ending at (1,1).

4. Properties of ROC curves

In any particular experiment when the observer is using a stable criterion, the two conditional probabilities define a point in the ROC space. This point is known as the operating point, and through it passes an experimentally undefined ROC curve that would be traced out if the observer varied his criterion. The slope of the ROC curve where it passes through the operating point is the likelihood ratio at criterion. It is an index of how cautious the observer is in accepting the event to be an orange. In standard detection theory usage, this index is called β; β can range from zero to infinity, being zero if the observer rejects as oranges only those events he is correctly certain are lemons, and infinity if he accepts as oranges only those events he is correctly certain are oranges.

The shape of the ROC curve depends on the structure of the two classes to be discriminated, and on the statistics of the noise which obscures the discrimination. If an orange event is frequently easy to identify correctly, the ROC curve will start from (0,0) with a long reach of high slope, while if the lemon event is often easily recognized, the ROC will reach (1,1) at the end of a stretch of low slope.

If ROC curves can have various shapes, it is apparent that one may cross another. It is then meaningful to talk as if discriminability had a unique measure? A theorem due to Green (1964) shows that it is. Green's theorem proves that for any yes–no ROC curve whatever, the area under the curve is equal to the maximum probability of a correct response in the conceptually matched two-alternative-forced-choice (2AFC) experiment. In the 2AFC experiment, the observer has to choose which of two observations, one being an orange, the other a lemon, was the orange. The probability of being correct in such an experiment seems intuitively a measure of discriminability, implying that discriminability is measured by any monotonic function of the area under the ROC curve.

The standard measure of discriminability, d', is not overtly related to the area under the ROC curve, yet has proved extremely useful. It derives from a specific type of discrimination, in which each likelihood ratio distribution is a monotonic transform of a Gaussian distribution, and the Gaussian distributions have equal variance. Using as unit measurement the common standard deviation of the underlying Gaussian distributions, d' is the separation of their means.

Symmetrical Gaussian ROC curves are completely specified by the associated value of d'. This means that their areas are known once d' is known; d' is therefore a monotonic transform of the area under the symmetrical Gaussian ROC curve, and a valid index of its discriminability. d' may be used as an index of discriminability for asymmetric ROC curves by equating the area of the asymmetric curve with that of a symmetric Gaussian curve. The d' of the symmetric curve then is an index of discriminability valid for the asymmetric curve.

5. ROC CURVES IN VIGILANCE

Oranges and lemons have been convenient labels for the two classes of events to be discriminated. The classes have been treated symmetrically. Why should an orange have preferential treatment over a lemon, or vice-versa? The situation changes when it comes to detecting a signal in a noisy background. The mathematical analysis is still symmetrical, but one does not think of the non-signal event in the same way as one thinks of the signal. The tendency is to make a non-symmetric analysis. The classical methods of psychophysics, in which subjects were chastized for false reports of signals, but not for false reports of non-signals, strikingly illustrates this asymmetry.

A signal normally is thought of as something added to the non-signal. The subject normally knows quite well what the non-signal would be, were it not obscured by noise. He does not know so well, however, what the signal would be without the noise. There is an essential asymmetry between the signal event class and the non-signal event class, that usually the subjects knows less about what is a valid example of the signal class than he does about what constitutes a non-signal. When this asymmetry does not occur, one usually finds that the tendency to analyse one class as signal and the other as non-signal also fails to occur, and the experiment is thought of as discrimination experiment rather than a detection experiment.

To see the effect of the asymmetry between the knowledge of the signal class and knowledge of the non-signal class, consider the detection of a 1000 Hz tone. Suppose that the subject is very precise in what he knows, so that the only uncertainty he is faced with is whether the tone, which he knows always to start from zero phase, initially goes positive or negative. This sign uncertainty is quite enough to demonstrate the main effects of asymmetry. The only noise which should affect the discrimination is noise which is like a 1000 Hz tone starting at zero phase, and initially going either positive or negative. Suppose the signal always starts by going positive, though the observer does not know that. There are four possible cases:

(1) The event is non-signal, and the noise starts positively.
(2) The event is non-signal and the noise starts negatively. In both these cases, the effect is to make the observation seem more like a signal.
(3) The event is a signal (starting positively) and the noise starts negatively. In this case, the observation is more like a non-signal.
(4) The event is a signal and the noise starts positively. Now the observation is less like a signal, but even less is it like a non-signal. It seems like an amplification of the signal, and will unfailingly be properly identified.

Of the four cases, none leads to easy and correct identification of the non-signal case, but one gives an easy identification of the signal. In terms of the ROC curve, easy identification of the signal leads to an initial portion of fairly steep slope, but with no easy identifications of the non-signal there will be no final portion of very shallow slope. The ROC curve is asymmetrical, or skewed.

The same analysis may be made, at least conceptually, for any detection situation. Always, if the subject knows less about what is a signal than about what is a non-signal, the ROC will be skewed in the same way. The curve will cling to the left edge of the ROC space longer than it does to the top. The less the observer knows about the exact characteristics of the signal, as compared to the non-signal, the more skewed will the curve be. Particularly in vigilance experiments, the asymmetry and skew will be marked, since the subject has much more opportunity to renew his acquaintance with the non-signal than he does with the signal.

Specific examples for an ideal observer of the skew due to uncertainty of signal specification are displayed in figs. 1 and 2. The symmetric ROC of fig. 1 represents possible detection behaviour for a signal of normalized energy 1.82 units in Gaussian noise of 1 unit power per unit bandwidth (cf. SWETS et al., 1961). The skewed ROC curve represents a signal of the same type in the same noise, but now the observer does not know in which

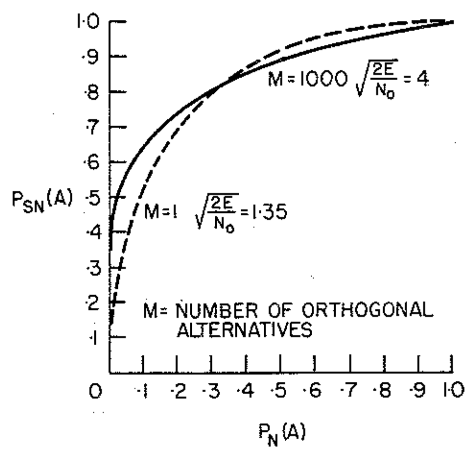

Fig. 1. ROC curves for a simple signal in Gaussian noise. The symmetrical curve is for the case where the observer knows the signal exactly, while the skewed curve represents the case in which he knows the signal exactly except that he does not know in which of 1000 intervals it may appear. The skewed curve is calculated for a signal with nearly nine times the energy required for the unskewed curve.

of 1000 different places the signal occurred. The energy required to obtain the same detectability is about 16 units with this much orthogonal specification uncertainty. While fig. 1 demonstrates the skew of the ROC curve and depression of detectability associated with orthogonal specification uncertainty, it also demonstrates that a lot of uncertainty is needed to produce a strong effect. However, if the uncertainty is not orthogonal, much less uncertainty may give quite dramatic effects. In the left panel of fig. 2, ROC curves are displayed for signals of two strengths. The uncertainty is in the sign of the signal and to some extent in its amplitude, which may be either 1/2 or 3/2 the nominal amplitude. Symmetric curves of the same detectability are displayed in the right panel for comparison. In spite of the fact that the uncertainty is limited to four possibilities for the signal instead of 1000, the skew in fig. 2 is more marked than in fig. 1. The important factor in determining the skew in an ROC curve is not the amount of uncertainty but its form.

6. Skewed ROC curve and the interpretation of data

There is one point in ROC space where a skewed ROC curve crosses its matched symmetrical Gaussian curve (see fig. 1). Tabulated values of d' assume that the ROC through the operating point is symmetrical Gaussian, so if the operating point is where the actual ROC crosses its matched Gaussian ROC, the tabulated value of d' will be correct. More commonly,

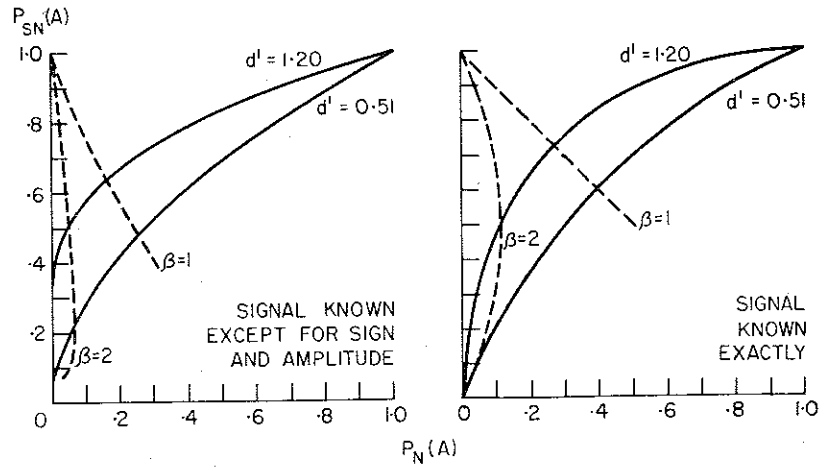

Fig. 2. ROC curves for a simple signal in Gaussian noise. The skewed curves represent signals for which the observer knows the signal exactly except for its sign and whether its amplitude is 1/2 or 3/2 the nominal amplitude. The two curves are for signals of different energy levels. The symmetrical curves are matched in area with the skewed curves. Isocriterion curves — lines connecting points where the ROC curves of a family have the same slope — are drawn for beta equalling one and two for both families, to illustrate the proposition that skewed ROC curves have at any point a lower slope than the symmetric Gaussian curve through the same point.

the operating point will be to the left of the cross-over, in which case, the tabulated value will belong to a symmetric Gaussian curve of too high detectability. In a vigilance experiment, this will be particularly true, both because the skew of the true ROC will be considerable, and because the disproportionate probability of a non-signal event demands that considerable evidence point to there having been a signal before the observer actually accepts the fact. The operating point in a vigilance experiment must therefore be in a region where the true ROC curve has quite high slope.

If the experimental conditions vary only the detectability of the signal, without influencing the criterion or the specification uncertainty, then perhaps tabulated values of d' may properly describe the results of the experiment. They will all be too high, but the absolute level of detectability is rarely very interesting in a vigilance context. If the experimental variations affect the criterion, then the tabulated values of d' will change, even though the true detectability of the signal may not.

Tabulated values of both indices presume that the ROC is symmetric Gaussian. At any particular point, the skewed ROC is always less steep than a Gaussian curve that crosses it at that point, so that tabulated values of β

are always too high. For the curves of fig. 2, tabulated values mostly run too high by a factor of between two and three in the region usually of interest. More realistically skewed ROC curves may give tabulated values in error by larger factors. This point may lie at the root of the dilemma encountered by JERISON et al. (1965) who found psychologically unrealistically high values of β in a vigilance experiment, and developed an ingenuous theory of mixed modes of response to account for them.

Skewed ROC curves change slope very fast near the left edge of the ROC space, and slowly near the right. When the appropriate operating point is in a region of steep slope, reasonably large variations in β cause small shifts of operating point in the ROC space. With slightly less caution demanded, reduction in β can mean significant rightward shifts along the ROC curve, while an equivalent increase in β means very little leftward movement. Observers differ, not only in β, but also in detection efficiency. If the operating point is in a region of not too high slope, it is quite conceivable that a slightly incautious or 'trigger-happy' poor detector might give a great number of false alarms, yet get fewer correct reports of signals than does a cautious good detector who gives almost no false alarms. It was largely this effect that dissuaded WIENER (personal communication, 1964) from using detection measures in his studies. He found very little correlation across observers between detection probability and false alarm rate, and concluded that the assumptions of detection theory were violated.

When the operating point is further from the 'knee' of the skewed ROC curve, trigger-happy observers should be less of a problem. Variations in detection ability, however, should then place the operating points for different observers along an isocriterion curve—the locus of points having the same criterion on different ROC curves of a single family. An isocriterion curve for skewed ROC curves does not follow an isocriterion curve for symmetrical ROC curves, as the examples in fig. 2 demonstrate. Hence, alterations in detectability without real changes in criterion will show up in the tabulated values as correlated changes in detectability and criterion. The same applies to changes in criterion without real changes in detectability. The expectation of spurious correlations, sometimes negative and sometimes positive, between values of d' and β taken from tables, suggests that one should avoid making any inference about behaviour from such correlations unless the obtained correlations can be shown not to be due to skew in the ROC curves. In particular, results from single operating points cannot be construed as giving evidence that good detectors tend to be either cautious or incautious.

7. SUMMARY

Consideration of the mechanics of detection from the point of view of subjective probability clarifies the usefulness of the ROC, even though the

ROC does not overtly involve subjective probability. In vigilance experiments, the ROC may be expected to show more or less severe skew, which affects the interpretation of the various indices of detection behaviour. The detection index, d', may give meaningful, though usually high values, and should be interpreted with care. The index of caution, β, will almost always be too high as read from tables, sometimes drastically so, and interpretations of β values must be regarded as dubious. Spurious correlations of β and d' are to be expected from the tabulated values, and without consideration of the ROC curve itself, no such correlation should be taken seriously.

The fact that these cautions are derived from consideration of the bases of detectability theory demonstrates that detection theory remains a potent instrument in the study of vigilance. A too restricted view of detection theory may lead to problems, but considerations of the data in generalized terms of detection theory is most likely to be helpful in understanding substantive problems.

REFERENCES

BROADBENT, D. E. and M. GREGORY, 1963. Brit. J. Psychol., **54**, 309—323.

BROADBENT, D. E. and M. GREGORY, 1965. Human Factors, **7**, 155—162.

EDWARDS, W., H. LINDEMANN and L. J. SAVAGE, 1963. Psychol. Bull., **70**, 193—242.

GREEN, D. M., 1964. Psychonomic Society Annual Scientific Meeting.

JERISON, H. J., R. M. PICKETT and H. H. STENSON, 1965. Human Factors, **7**, 107—127.

MACKWORTH, J. F. and M. M. TAYLOR, 1963. Canad. J. Psychol., **17**, 302—325.

SWETS, J. A., W. P. TANNER and T. G. BIRDSALL, 1961. Psychol. Rev., **68**, 301—340.

WIENER, E. L., G. K. POOCK and M. STEELE, 1964. Percept. Motor Skills, **19**, 435—440.