Applied Ergonomics 1971, 2.2, 74–78

The analysis of industrial inspection

J. J. Sheehan

Department of Industrial Engineering, University of Massachusetts

and C. G. Drury

Ergonomics Section, Pilkington Brothers Ltd, St. Helens, Lancs

Inspection is a widespread activity in industry with many problems. This article describes one of the methods, for analysing human inspection performance, which has been developed and applied recently, with some success.

Typically an industrial inspector's job is to divide a batch of product into acceptable items and rejectable items, where 'acceptable' and 'rejectable' are defined in terms of standards agreed between the manufacturer and the customer. This he does either by examining each item in turn for defects or by examining a group of similar items in his visual field and noting those which do not conform (eg Fox & Haslegrave, 1969).

If the question is asked 'how good is this inspector?' the answer is usually given by the percentage of rejectable items which he correctly detects. This ignores the rest of his job, which is correctly to accept all acceptable items. Thus two types of error can occur:

Type 1 error : classifying acceptable parts as rejectable
Type 2 error : classifying rejectable parts as acceptable.
For convenience a Type 1 error is often called a 'false alarm' and a Type 2 error is called a 'missed signal.' The 'signal' in this case is the particular defect which renders the item rejectable (not necessarily the same defect for all rejectable items, of course).

This paper describes a method for combining information on these two types of error using the 'Theory of Signal Detection,' a technique first applied in laboratory experiments about 15 years ago (Tanner & Swets, 1954) and now beginning to be applied to industrial inspectors (Wallack & Adams, 1969). The method is first described briefly and then applied to a case study of industrial inspection, described in more detail elsewhere (Drury & Sheehan, 1969), to show the usefulness and limitations of the theory.

Theory of signal detection applied to inspectors

The inspector usually is presented with a visual field, most of which is 'good' product and some of which is 'faulty.' However, the 'good' items are not perceived as featureless, and therefore obviously differentiated from the bad items with defects, for two reasons. Firstly dust and just-acceptable small irregularities are present. Secondly the nervous system, translating the light pattern falling onto the eye into a series of electrical impulses, is subject to random electrical impulses. Both of these, the dust and irregularities and the random nerve impulses, can be regarded as 'visual noise,' which is superimposed on both the good items and the faults.

Considerable success has been achieved in recent years by treating the human operator, in a signal detection system, as if he were making a statistical decision of whether a 'signal' is present or whether 'noise' alone is present. The work described in Swets (1964) shows how such a treatment is an accurate description of human performance in a variety of experimental signal detection tasks. A diagram of the decision situation is shown in Fig 1. It is assumed that the human operator can assess the amount of activity generated in his nervous system by the particular item being observed. This amount of activity is not a single value, but varies to some extent even during the viewing of one particular item and more so during the course of viewing a batch of items. The variation in activity level can be represented by the normal probability distribution. If there is no defect then the probability distribution is that of noise alone: the left hand one. If an item with a defect occurs giving a signal of a particular strength, there will still be noise added to it and the distribution will shift to the right as shown by the 'Signal + Noise' distribution. The separation of the two distributions depends on the strength of the signal, and the detectability of the signal is the ratio of this separation to the standard deviation of the distributions. This parameter is usually symbolised as d^1.

It is apparent that the two distributions overlap to a certain extent; if the random variation of noise reaches a

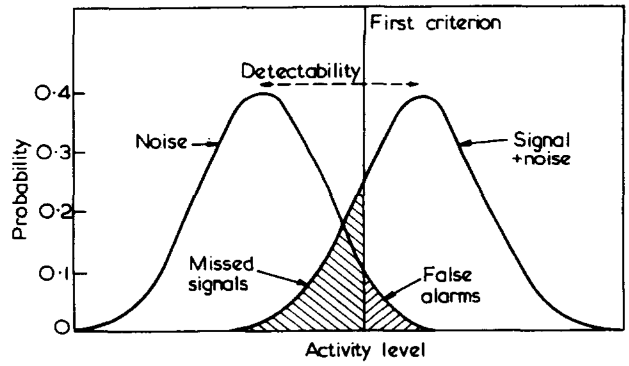

Fig 1 Decision situation, first criterion.

large activity value or the random variation of signal plus noise reaches a small activity value, then the amount of nervous activity will be the same. The operator must 'draw a line' somewhere, ie choose a criterion level such that any level of activity above the criterion will cause him to say 'signal present' and reject the item as faulty. In Fig 1 he has chosen a high criterion and so has few false alarms, that is he rarely rejects good items. The price he pays, of course, is that he misses a fair proportion of bad items. Fig 2 shows a low criterion where the situation is reversed: he detects more bad items but rejects more good items.

In practice the human operator can choose from a continuous range of criteria between these two extremes. Each criterion line will, by the point at which it cuts the two distribution curves, define a pair of values: probability of correct detection (on the signal + noise curve) and probability of false alarm (on the noise curve). For a constant value of detectability d^1 a curve of one against the other can be plotted. A series of such curves is shown in Fig 3. It is obvious that the two parameters, d^1 and the criterion line position, can be varied independently. Thus for a fixed value of d^1 we can choose an economic trade off between correct detections and false alarms. The only way to improve both of these factors simultaneously is to increase d^1, by improving the signal to noise ratio. However, a doubling of d^1 requires four times the signal to noise energy ratio, so that improvements in this direction are difficult.

Thus, it is clear that d^1 is a measure of the difficulty of the task and is largely outside the control of the inspector. It can only be improved by altering the task conditions. The criterion line position, on the other hand, may be chosen and varied explicitly or implicitly by the inspector, and is often influenced by several factors which may bring pressure upon him. These aspects are considered further in the discussion section.

The implied assumptions in this theory are that the variances (the square of the standard deviation) of the two distributions are equal and that both are normal probability distributions. Although these assumptions are unlikely to be satisfied fully in practice, they produce a theory which fits the observed facts in numerous experimental situations.

The case study of inspection

A factory in the North Eastern United States produces small metal parts, which we shall call 'hooks' to preserve the anonymity of the company. The company has a reputation for quality: its competitors produce what is considered an inferior product at a lower price. The quality requirements on the hooks are stringent and likely to become more stringent. Throughout manufacture Statistical Quality Control schemes are used to control dimensional accuracy but a final 100% visual inspection is still required. Even this is subject to a sampling check inspection by 'oversight' examiners.

The hooks studied here were not sold directly to the customer but cast into blocks of about 30 hooks in an assembly department of the factory. This gave an 'in house customer' so that the efficiency of the final inspection system could be evaluated.

A diagram of the hook is shown in Fig 4. An analysis of the 32 defect types found at final assembly showed that 81·5% of all scrap (costing $13 000 p.a.) was caused by only six defects. These ranged from marked or twisted points to missing levers and mixed types of hooks.

Method of investigation

In order to determine the effectiveness of inspection, six inspectors were selected and given twelve different batches of hooks to examine. The inspectors were selected to cover

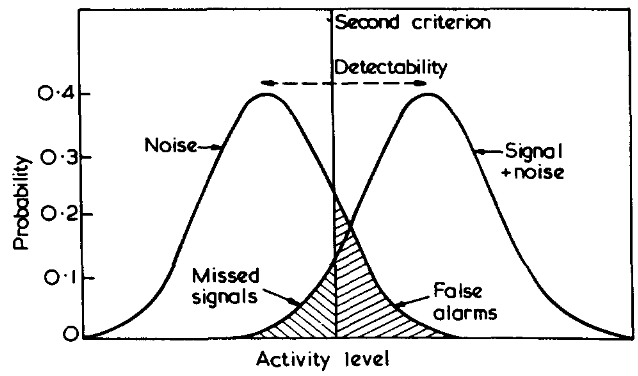

Fig 2 Decision situation, second criterion.

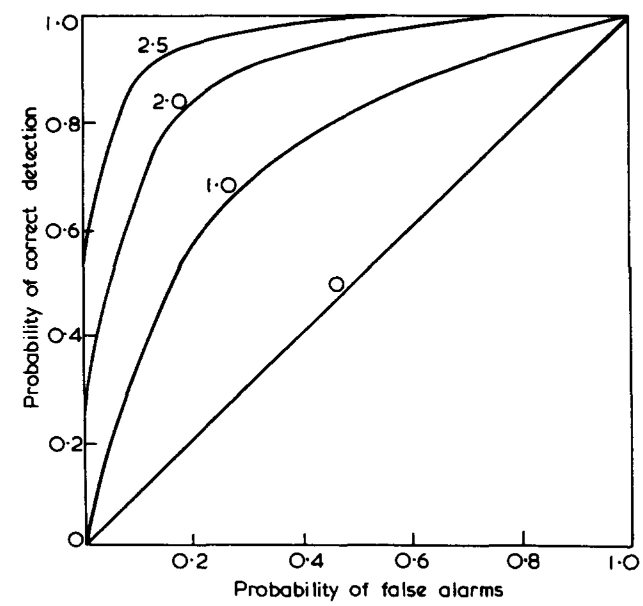

Fig 3 Performance curves on a signal detection task, (parameter of curves is detectability).

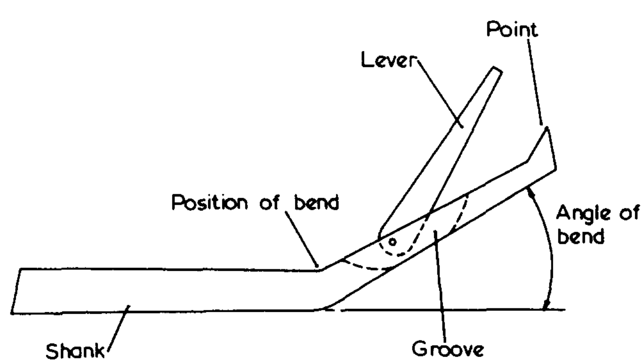

Fig 4 Hook: overall length 30·48 mm (1·2 in), thickness 0·3048 mm (0·012 in).

a range of ages (30 to 64 years), a range of visual acuities (measured using a Landolt ring test) and of approximately the same ability at inspection (measured by their Merit Ratings). The twelve batches of hooks comprised:

6 batches of 200 good hooks and 50 with a single defect
5 batches of 200 good hooks and 25 of each with two defects
1 batch of 250 good hooks

Each inspector inspected six batches per day over four days. During that time she had inspected each of the twelve batches twice. Randomisation was used to remove bias due to time of day and learning effects.

Four different illumination levels were used, 1937, 968, and 646 lx (180, 90 and 60 foot-candles) on the first three days and an illumination level of the subject's own choosing on the fourth day. The working conditions were otherwise normal, using the inspectors own workplace and the company time standard of 15 minutes per batch. Inspection times were measured in case there was a trade-off between time and accuracy but none was observed, so that the same inspection time can be assumed for all conditions.

As a result of the analysis of this experiment, in which the subjects did not know which defects to expect, a second, smaller experiment was undertaken using the same inspectors but telling them before each trial which defect would be present. Here the six 'single defect' batches were used and each inspector inspected each batch once only.

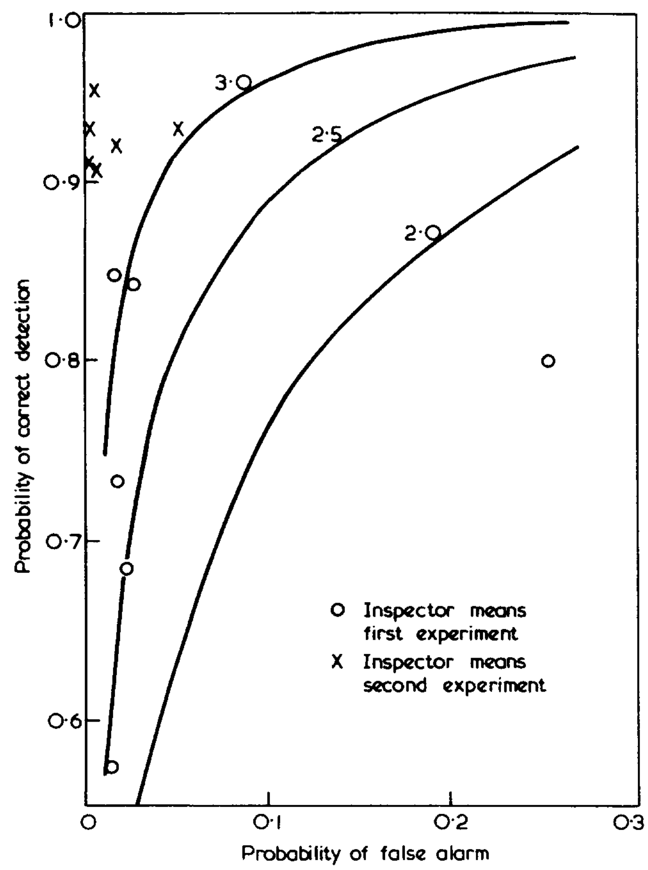

Fig 5 Inspectors' mean performances.

Results

Previous analysis (Drury & Sheehan, 1969) had shown that there were no effects attributable to learning, illumination level or visual acuity. Signal detection analysis confirmed these findings. The method of analysis used was to calculate the probabilities of Type 1 and Type 2 errors and use tables (Swets, 1964) to obtain the value of d^1, the detectability of the defect for the particular inspector. However, in some cases the observed probabilities or error were zero, indicating an infinite value of d^1. Other workers in this field have circumvented this difficulty by calling a frequency of zero of a particular error 'less than ½ an error' and thus replacing each zero by ½ to obtain a lower bound of d^1. Here it was decided to average the results for each inspector or for each defect and obtain average values of the probabilities and 'average' values of d^1. This effectively ignores interactions between inspectors and defects. It was felt that this decrease in precision would more than compensate for the arbitrary assumption that zero errors means less than ½ an error.

Table 1 shows in three columns the average values of the probability of correct detection (= 1 − p(Type 2)), the probability of false alarm (p(Type 1)) and the corresponding values of d^1, for the inspectors and defects in Experiment 1 and for the inspectors in Experiment 2. Fig 5 shows the inspector's averages plotted onto a diagram similar to Fig 3 with the lines of constant d^1 superimposed.

Both Table 1 and Fig 5 show that one inspector (No. 4) has a much worse performance than the other inspectors in both experiments. When she was questioned after the experiments, this was found to be due to her rejecting large numbers of acceptable hooks because of a surface blemish on the shank. This was not considered a defect as it in no way affected the performance of the hook and would be covered during the subsequent casting process. This illustrates the usefulness of the Signal Detection approach. Inspector No. 4 had a probability of correct detection similar to all the other inspectors; however, her false alarm probability was much higher. Measuring both aspects of performance simultaneously was necessary to alert the experimenters and management to the mistake, which could have been particularly expensive to the company. Since she was detecting a completely different and irrelevant fault, her results were not included in the further analyses.

It is of interest to see whether the performance of the inspectors is different when defects occur singly or in pairs, and when the inspectors know what type of defect to expect. Comparing the figures for single and paired defects, in Table 1, suggests there is no real difference, and a statistical test (Student's 't') on the values of detectability for single and paired defects confirms no significant difference. This was true whether or not subject 4 was excluded from the analysis. A similar comparison and significance test of the detectability values for the same inspectors in the two experiments showed a significant improvement ($p < 0·01$) in the second experiment, where the defect type was known in advance.

A point of practical importance, shown in Fig 5, is that all of the inspectors' results are to the left of the figure. This means that their criterion is of the sort shown in Fig 1. They are not making many false alarms, but they are missing quite a number of defective items. They should

Table 1 Values of probabilities and detectabilities in each experiment.

Experiment	Classification	p (correct detection)	p (false alarm)	detectability (d^1)
1	Subjects 1	0·85	0·01	3·3
	2	0·73	0·02	2·7
	3	0·84	0·02	3·0
	4	0·80	0·24	1·5
	5	0·57	0·01	2·3
	6	0·69	0·03	2·4
1	Defects AB	0·81	0·02	3·1
	BC	0·86	0·02	3·1
	CD	0·71	0·01	2·9
	DE	0·77	0·007	3·1
	EF	0·78	0·02	2·9
	A	0·68	0·01	2·7
	B	0·92	0·01	3·6
	C	0·76	0·04	2·5
	D	0·87	0·008	3·5
	E	0·98	0·008	4·4
	F	0·69	0·04	2·3
2	Subjects 1	0·93	0·003	4·2
	2	0·95	0·008	4·1
	3	0·91	0·003	4·0
	4	0·93	0·05	3·1
	5	0·90	0·005	3·8
	6	0·92	0·02	3·5

detect many more defective items if their criterion line were moved to the left, to a position similar to Fig 2. This would be equivalent, in Fig 5, to moving their criterion point along the d^1 curve up and to the right of the figure. They would necessarily reject more acceptable hooks, but they are at present behaving as if they assume this is much more costly than to accept a faulty hook. Because the hooks are subsequently cast in batches of 30 or so, all of which must be perfect, their criterion is based on the wrong assumption. In fact, every faulty hook accepted will subsequently lead to a batch of 30 being rejected, which is clearly more costly. Therefore to move their criterion; so as to improve the probability of rejecting the faulty hooks, ie to increase the correct detections even if more good hooks are rejected also, would clearly be the better strategy.

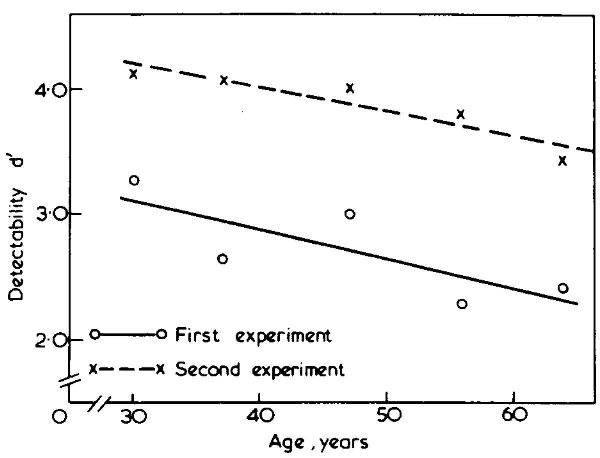

Fig 6 Effect of age on detectability.

An interesting indication in the data is the effect of the inspector's age on inspection performance. It was found previously (Drury and Sheehan op cit) that there was a mild correlation between increasing age and decreasing probability of correct detection. However, when the detectability values are plotted against age (Fig 6) a clearer picture emerges. Both experiments show a decline of detectability with age. When regression lines were fitted to each experiment, the results (Table 2) are just statistically significant for experiment 2. Both experiments show a very similar change in detectability with age, a decrease of approximately 0·2 in d^1 for each 10 year increase in age. Although this may be of practical importance, with only 5 subjects the data should be regarded as suggestive rather than conclusive.

Discussion and implication of the results

These results have been analysed both conventionally (Drury and Sheehan op cit) and now by 'Signal Detection Theory' analysis, so that the advantages and drawbacks of

Table 2 Correlation of detectability with age

	Experiment 1	Experiment 2
Correlation coefficient	−0·78	−0·92
Significance	p = > 0·05	p = <0·05
Slope of regression line	−0·023	−0·018
Intercept of regression line	3·8	4·8

this later analysis can be seen. The major advantage of analysis by Signal Detection Theory is that it gives two measures, the detectability and the criterion, which are more useful than either of the probabilities from which they are calculated.

Detectability is an unbiased measure of the difficulty of the inspection task to the inspector. Here, for example, it has been used to quantify the age effect far more precisely than was possible using either probability of error alone. The effects of the other variables important in inspection problems, such as fault density, time allowed for inspection, size of defect and length of working period, could be evaluated in this way using a measure of difficulty independent of the type of fault or even the sense used to perceive it. This would both enable better inspection systems to be designed and enable meaningful comparisons between different inspection tasks to be made.

The criterion measures the combined effect of the various pressures on the inspector. In laboratory tasks these are monetary cost and values, but there are more than monetary pressures on an industrial inspector. There are the conflicting demands of production, quality control and sales, as well as personal and group pressures, to be added to any direct financial inducements. The criterion measures the resultant of these forces.

With these two measures, the options open to management are more clearly defined. To move along one of the curves in Fig 3 involves only changing the criterion. This can be done, for example, by rapid feedback of information to the inspector where his present criterion is compared with where management would like it to be. An inspectors' current criterion could be described to him as 'too stringent' or, as was the case with this data, 'not stringent enough.' Such a change of criterion involves very little change in the situation but it achieves only a better balance between the two errors.

A change in detectability (moving up to a higher curve in Fig 3) requires more management effort but has potentially greater rewards. In experiment 2 the inspectors received prior information about which faults a batch would contain. This effectively improved the detectability of the faults and reduced both errors simultaneously. Other ways of improving the detectability are better training, improved lighting (if the inspection is visual), and generally the application of good ergonomic practice to the design of the job and the environment.

The main drawback of this analysis was that it did not find a very interesting, and economically important, effect found by close scrutiny of the raw probability data. This was an all-or-none effect in performance of the first experiment. It appeared from the original analysis that the inspectors had a tendency to find almost all of the defects, of a particular type in a batch, or almost none of them. They behaved as if, having discovered a defect early in a batch, they were alerted to it and picked up almost all subsequent defects of that type. Failure to spot a defect early in a batch resulted in very few of this defect being found in the rest of the batch. The practical cure for this is to provide information for the inspectors about which defects are likely in particular batches, as was done in experiment 2. However, the superiority of performance in experiment 2 was clearly shown by the Signal Detection Theory analysis, and the same action recommended.

Recommendations

As a result of this study the following recommendations were made to the company.

1. Regular recalibration of inspectors. The experience of the inspector who consistently rejected hooks, because of a defect which was of no operational consequence, caused the management to institute a regular test, such as that used in these experiments, for all the examiners. This test is quick and simple, and disputed decisions can be resolved immediately between the inspector and her supervisor.

2. The Signal Detection analysis suggests that a new look should be taken at the criterion chosen by the inspector. The economic consequences of the two types of error should be assessed, and used to calculate an optimum criterion to minimise the total cost. This criterion can thus be taught to the inspectors by the recalibration sessions, coupled with feedback of knowledge of their current performance in relation to the desired standard.

3. Feed-forward of fault data to the inspectors. The large improvement in detectability when the fault type is known in advance shows that an attempt should be made to provide prior knowledge of fault types. The present oversight inspectors should perform a sampling on the batches coming to the inspectors, to provide information for the inspectors on the fault types to be expected in each batch. Using this procedure in the experiments has reduced the missed defects from 17½% to 7½%. This represents a large saving to the company by reducing subsequent processing of defective items.

After two years operation of this revised inspection system the company, although not wishing to disclose financial results, has found that it is working very well.

Acknowledgments

The authors would like to express their appreciation for the assistance given by Professor R. W. Trueswell and the Department of Industrial Engineering, University of Massachusetts, where this research was undertaken.

References

Drury, C. G., and Sheehan, J. J.
1969 *International Journal of Production Research,* 7, 333–341. Ergonomic and economic factors in an industrial inspection task.

Fox, J. G., and Haslegrave, C. M.
1969 *Ergonomics,* **12**.5, 713–721. Industrial inspection efficiency and the probability of a defect occurring.

Tanner, W. P., and Swets, J. A.
1954 *Psychol. Rev.,* **61**.6, 401–409. A decision-making theory of signal detection.

Swets, J. A.
1964 'Signal detection and recognition by human observers.' New York: Wiley.

Wallack, P. M., and Adams, S. K.
1969 *American Institution of Industrial Engineers Transactions,* **1**.1, 33–44. The utility of signal-detection theory in the analysis of industrial inspectoı accuracy.