HUMAN FACTORS, 1987, 29(6), 645–652

Signal Detection Theory and Probability Matching Apply to Vigilance

ANGUS CRAIG,[1] *University of Sussex, Brighton, England*

Many vigilance tasks require discrimination of infrequent signal events from frequent nonsignals. During performance on these tasks signal detections often decline. But this does not generally signify a loss of vigilance, if one can rely on signal detection analyses showing that signal discriminability remains constant during a vigil and hence that neither attention nor signal processing has waned. This paper confirms that signal detection theory does provide a good fit to vigilance data and that the analyses can therefore be relied on. The paper also shows that probability matching (of signal reports to signal occurrence) occurs. In the main it is the adjustment of report rate toward matching by an alert observer, in control of his or her performance, that produces the vigilance decrement in detections.

INTRODUCTION

By tradition, vigilance involves sustained monitoring for an infrequent signal event. The signals may be transient, discrete events or they may consist of continuous movements away from a norm. They can occur against a background of continuous noise or may be embedded in a train of temporally or spatially discrete nonsignal events. It is the single most characteristic feature of behavior during vigilance that the proportion of signals detected declines significantly as time progresses.

The importance of these failures of vigilance is that they seem to give a negative answer to the question, Can we sustain attention? This conflicts with earlier findings summarized by Thorndike (1926) indicating that one can apparently concentrate for a long time on a single task, sustaining attention to it for periods of up to two hours, without any serious falloff in performance. But do the vigilance data really contradict these earlier findings? This question has to be reexamined because the results of applying signal detection theory (SDT) to vigilance imply that in most cases the ability to discriminate between signals and their background can be sustained and that the changes in behavior that are found reflect active processes of adjustment as observers get used to the low likelihood of signal occurrence that is peculiar to the vigilance task (Craig, 1978; Vickers, Leary, and Barnes, 1977; Williges, 1976). The exceptional cases in which the observer does not maintain the ability to make the necessary discrimination seem to involve tasks in which the workload is high, the demands for repetitive, effortful resource allocation being too great to be met indefinitely (Fisk and Schneider, 1981; Parasuraman, 1979). These current interpretations of detec-

[1] Requests for reprints should be sent to Angus Craig, MRC Perceptual and Cognitive Performance Unit, University of Sussex, Brighton, BN1 9QG, Sussex, England.

© 1987, The Human Factors Society, Inc. All rights reserved.

Downloaded from hfs.sagepub.com at PENNSYLVANIA STATE UNIV on September 11, 2016

tion decrements rely heavily on the applicability of signal detection theory to vigilance data.

APPLYING SIGNAL DETECTION THEORY

The proportion of signals detected in vigilance may be low because the observer is incapable of accurately discriminating between signal and nonsignal events or because the observer's responses are biased in favor of nonsignal occurrence, demanding little evidence to judge that an event is a nonsignal but requiring strong evidence before reporting that a signal occurred.

Reports of signal occurrence that are correct (hits) and those that are incorrect (false alarms) are jointly determined by the observer's capacity for discriminating between signal and nonsignal and by his or her bias toward reporting signal occurrence. Even when capacity remains fixed hits and false alarms will vary over a wide range as response bias changes. In SDT the relative operating characteristic (ROC) function, which defines the relation between the proportion of hits and the proportion of false alarms as response bias varies, is assumed to be linear on a binormal graph (normal deviate or z-score axes) of the form

$$Z(Hit) = B.Z\,(False\ Alarm) - A \qquad [1]$$

where $Z(X)$ is the normal deviate corresponding to the proportion $P(X)$. The slope, B, reflects the ratio of the variance of sensory effects that arise from nonsignal occurrences to the variance of effects from signals. These properties of the ROC are illustrated in Figure 1.

Discriminability is indexed by a measure of the distance that the ROC lies from the chance line (where hit and false alarm proportions are equal). In earlier studies applying SDT to vigilance it was commonly assumed that the ROC took the specific form

$$Z(Hit) = Z(False\ Alarm) - A \qquad [2]$$

in which the line has a slope of 1.0 (signifying equal variance of signal and nonsignal effects, as in the central image in Figure 1) and A is the familiar discriminability index, d'. For this model the curve of P(Hit) against P(False Alarm)—the curvilinear ROC form—has a monotonically decreasing slope.

Taylor (1967) and Jerison, Pickett, and Stenson (1965) have argued against the use of this specific model in vigilance. Their arguments rest on hypotheses, not on empirical findings. Taylor argues plausibly that observers know less about what constitutes a valid example of a signal than they do about what constitutes a nonsignal (i.e., effects of signals will be more variable) and that this essential asymmetry will be reflected in the ROC, which will have a flat slope (less than 1.0) on the binormal graph as in Figure 1a. This will be especially so in vigilance because observers are much better acquainted with the more frequently occurring nonsignals. Jerison and his colleagues argued just as plausibly that the SDT measure of discriminability was valid only under alert observing and that this probably declined during vigilance, being increasingly replaced by less

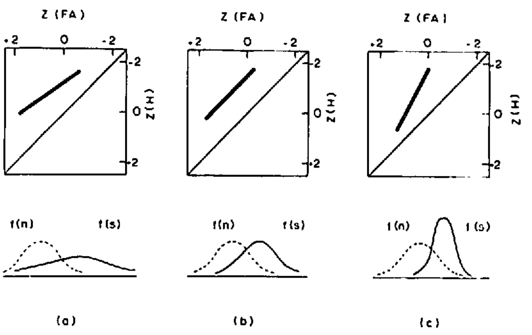

Figure 1. *The relation between the slope of the linear relative operating characteristic on a binormal graph and the variances of the underlying sensory effects arising from signal [f(s)] and nonsignal [f(n)] presentations.*

Downloaded from hfs.sagepub.com at PENNSYLVANIA STATE UNIV on September 11, 2016

alert observing or by distraction. If overt responses are made only to report signal occurrence, the effect of blurred observing and distraction will be an artificial flattening of the ROC slope; thus, as for Taylor, the expected ROC form is similar to the one depicted in Figure 1a. Taylor and Jerison agreed that these departures from the equal variance model would render suspect any interpretations that were based on it, a view that has been echoed by several investigators, including myself (Craig, 1977; see also, for example, Davies and Parasuraman, 1982; Long and Waag, 1981).

It is tempting to dismiss these criticisms, to point out that detection theory is not confined to the equal variance model and that vigilance researchers such as Nuechterlein, Parasuraman, and Jiang (1983) and Williams (1986) have reported that the general linear model of Equation 1, in which slope is a free parameter, acceptably fits the ROCs of most of their subjects. However, evidence either against or in support of the arguments of Taylor and Jerison has never been adequately examined. An attempt to do so is made in this section, which looks at the following questions: Do ROCs obtained in vigilance have particularly flat slopes? And do these slopes decline with time on task? The inquiry is admittedly limited because of restricted data access, as only data obtained by the author or his colleagues have been examined.

Craig (1977) presented evidence on the form of ROCs obtained in four vigilance studies that had been conducted by W. P. Colquhoun. Three studies used an auditory task; the fourth used a visual task. Events were discrete and transient, and signal probabilities were in the range of 0.05–0.20. The unifying feature of the studies was that in each case observers responded only when they thought a signal had occurred; they responded by means of four-point rating scales that permit assessment of the ROC form. A total of 258 individual data sets were examined. Of these, 191 (74%) contained entries in all cells of the rating response matrix, so that ROCs could be defined. Less than half (87, or 46%) of the ROCs had the monotonically decreasing slope that is implied by the equal variance model, and a significant number (57, or 30%) had a form that seemed more compatible with a steeply inclined ROC (slope > 1.0).

The data were also analyzed using the SDT curve-fitting program of my colleague Byron Morgan (Grey and Morgan, 1972). The program estimated not only the parameters A and B of Equation 1 but also the χ^2 goodness of fit of the linear model to the data. Estimates of the ROC slopes were obtained from 137 cases for which χ^2 lay between the 10% and 90% points, so that there was certainly no reason to doubt the linear model (goodness-of-fit criterion given by Fisher, 1970, p. 80). Of the slopes, 67% exceeded 1.0, and the mean slope was 1.41 (see Luce and Green, 1974, note to p. 312). The distribution of slopes is shown in Figure 2. (The positive skew is substantially reduced if the logarithm of the slope is used to scale the abscissa.)

Craig (1977) pointed out that ROCs with steep slopes had been obtained by Green and Luce (1973) in a signal detection task in which a response-time deadline was imposed on signal trials but not on nonsignal trials, a circumstance resembling the response condition in the studies by Colquhoun from which Craig's sample was drawn.

Recently I have examined the individual data sets from a series of studies reported in *Human Factors* (Craig, 1979, 1980, 1981) and from the first of two experiments reported by Craig and Colquhoun (1977). In these studies discrete transient visual events were dis-

Downloaded from hfs.sagepub.com at PENNSYLVANIA STATE UNIV on September 11, 2016

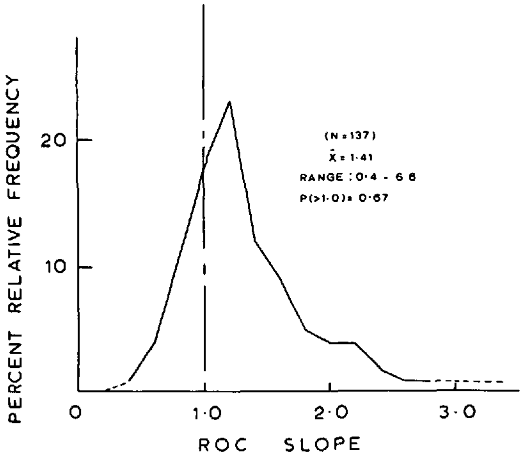

Figure 2. *Distribution of ROC slope estimates from data obtained in vigilance tasks in which a response was made only when a signal was thought to have occurred.*

played; signal probabilities were in the range of 0.05–0.15; and responses (on a four-point rating scale) were required to nonsignal as well as to signal events. In 121 of 248 individual data sets (49%) the ROC curve form could be defined (no zero entries in the cells of the response matrix); and of these ROCs a significant majority of 95 (79%) had the monotonically decreasing slope implied by the equal variance model. The latter proportion did not differ significantly from the proportion observed when the same events were presented in a discrimination task with 50% signals and with rest pauses between blocks of trials. In other words, when observers in a vigilance task distribute their responses across a rating scale in such a way that an ROC can be defined, the form of the ROC does not differ from that obtained in a signal discrimination task. (It is worth noting, however, that in the vigilance condition there was a disproportionately high incidence of failures to make full use of the rating scale—as though observers were simplifying the task by opting not to use the full rating scale but to respond in a simple yes–no manner).

The vigilance data were further analyzed using a variant of Dorfman's SDT curve-fitting program (see Swets and Pickett, 1982), which is similar to the program of Grey and Morgan (1972) referred to earlier. With the same goodness-of-fit criterion as before, this analysis revealed that the linear ROC defined in Equation 1 provided a satisfactory fit to 114 of the individual ROCs. The mean slope of these was 0.9511; average slopes for consecutive quarter-hour periods of the one-hour sessions were 0.9884, 0.8766, 0.9763, and 0.9674. There is no evidence in these data of any change in slope with time on task, or of any significant departure from the equal variance slope of 1.0.

To have to rely on data from one's own studies is not entirely satisfactory. It is therefore comforting to note that in a recent study with a visual vigilance task, Williams (1986) has obtained mean ROC slopes that approximate 1.0 in consecutive half-hour periods of a one-hour session.

The data considered in this section do not support the arguments of Taylor or of Jerison. Instead, they justify confidence in even the parametric SDT indices such as d', when estimated from vigilance data, although the recommended indices still remain the less restrictive ones, such as the measures of the area under the ROC ($P(A)$, $A(Z)$; see Swets and Pickett, 1982).

The apparent stability of the ROC slope within the vigilance session is important not only because it signifies an aspect of behavior that does not change during the course of a vigil but also because changes in slope would confound interpretations of discriminability deficits or criterion shifts.

CRITERION SHIFTS AND PROBABILITY MATCHING

The common observation that it is a criterion shift that produces the detection decrement leaves unexplained the reasons for the

Downloaded from hfs.sagepub.com at PENNSYLVANIA STATE UNIV on September 11, 2016

shift in criterion. There are two radically different views on why the criterion shifts: one view attributes the shift to a negative influence and the other to a positive one.

The negative view may be seen as a derivative of the expectancy theory of vigilance (Baker, 1963; see also Broadbent, 1971). On occasions, perhaps because expectancy is already low, the observer fails to detect the signal when it occurs. This failure causes a downward revision in the expectancy that a signal will occur and a concomitant adjustment of criterion. As a result of this lowered expectancy the observer will be even less inclined to detect the occurrence of the next signal event. And so begins the "vicious spiral" that produces the vigilance decrement.

The positive view, associated most notably with Craig, Vickers, and Williges (see, for example, Craig, 1978; Vickers, 1979; Williges, 1976), is that the criterion shift is the outcome of an active effort by the observer to satisfy some criterion—for example, to minimize errors, to balance errors of omission against those of commission, to reduce discrepancies between current and historic response rates, or to maintain a "criterial" cutoff in the middle of the sensory range as experienced on the task. Craig and Vickers predict that the adjustment process will stabilize where response and signal rates are equal—that is, where "probability matching" is attained (this will be a dynamic equilibrium and hence the variance in report rate will probably exceed binomial variance). Williges held that the criterion would be shifted toward its optimum (to minimize errors or maximize net gain). It is generally acknowledged that the obtained position of the criterion is less extreme than the optimum, an observation that is consistent with probability matching. A particularly clear example of response rate stabilizing at the matching level is shown in Figure 3. (Appropriately, the data are derived from one of Mike Loeb's studies.)

In order that within-session criterion adjustment be able to reduce the report rate toward the matching level, the report rate needs initially to have exceeded the signal rate. This means that over the session as a whole signal reports should occur more frequently than do signals. Craig (1978) showed that in the mean data in published studies, reports usually did occur more frequently than signals. This applies to individual report rates as well (at least as found in the author's own studies). Of the 248 individual data sets collated from the studies reported by Craig (1979, 1980, 1981) and from the first experiment reported by Craig and Colquhoun (1977), only 12 (less than 5% of the sample) report signals significantly less frequently than they actually occur during the first quarter of a session and 36 (less than 15%) during the final quarter. In that final quarter 158 (64%) show approximate probability matching.

As Vickers has emphasized (Vickers, 1979; Vickers and Leary, 1983; Vickers, Leary, and Barnes, 1977), exposure to the statistical structure of the task begins at the earliest stages of training—not merely at the start of the test session. Training sessions often include task samples at the low signal probability that will be encountered later in the test session, explaining why some partial adjustment is often seen to have already occurred by the start of the test session (Vickers, Leary, and Barnes, 1977).

It is a critical feature of the positive view of criterion adjustment that it should apply as easily to increments in report rate as to decrements. Referring again to my own collated data, I find that report rate changed significantly ($p < 0.05$, binomial test) between the beginning and end of the test session in 158 individual cases. In 25% of these cases the report rate actually increased over the session,

Downloaded from hfs.sagepub.com at PENNSYLVANIA STATE UNIV on September 11, 2016

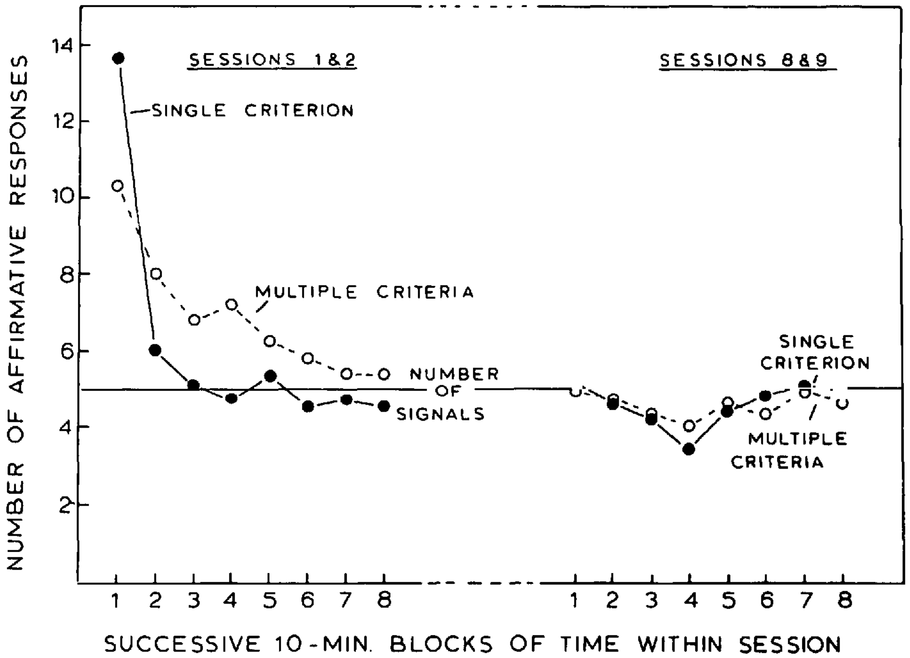

Figure 3. *Mean number of responses in successive 10-min periods of the first two and final two daily sessions of a vigilance task that was performed for nine consecutive days by groups employing single or multiple response criteria. The horizontal line shows the number of signals per 10-min period. Data from Binford and Loeb (1966); from Craig (1983).*

whereas in 75% it declined. In either direction 70% of the changes brought the report rate closer to the probability-matching level. The observation made by Colquhoun and Baddeley (1967) that β actually declined during the session when signal probability was increased from 2% during training to 18% during testing is consistent with this positive view of criterion adjustment. Vickers has indicated that these effects may be enhanced by an active hunting process when signal probability is suddenly stepped up or down: when the probability is reduced, the observers' initial reaction is to reduce β, as if to hunt for the signals they expect to meet; when the probability is increased, they rapidly increase β, as though suddenly finding themselves reporting too frequently (Vickers, 1979; Vickers, Leary, and Barnes, 1977).

Adjustment toward probability matching will depend on discriminability. The less physical separation there is between signal and nonsignal events, or the less able the observer is to resolve the difference, the less adjustment there will be, so that in vigilance, with its usual low signal probability, the higher the response rate will remain; whereas the greater the discriminability, the greater will be the degree of adjustment and hence the lower the report rate. Craig's data (1979, 1980, 1981) support this: for each group in each study there is a significant negative correlation between report rate and A', the area under the ROC (over the studies, the combined correlation using Fisher's technique is: $r = -0.730$, $df = 72$, $p < 0.001$). Note that this result is at variance with the more traditional negative view about criterion adjustment, according to which there should be a greater decline when signals are

Downloaded from hfs.sagepub.com at PENNSYLVANIA STATE UNIV on September 11, 2016

not easy to discriminate than when they are. The result also clashes with the arguments put forward by Jerison, Pickett, and Stenson (1965) to explain why the correlation between detectability and report rate should generally have a positive sign—or, in their terms, why d' and β should be negatively related (despite their own observations that d' and β tended to be positively associated). The arguments presented by Taylor (1967) referred to in the previous section also imply a negative correlation between d' and β and are inconsistent with the evidence. (It is also of interest to note here that Long and Waag [1981], who present five sets of data, four of them moderately close to probability matching, consistently found a positive correlation between d' and β [for the less probable signal event] but chose to regard the correlations as spurious.)

The positive link between d' and β is illustrated in Figure 4, based on the estimates obtained from the data of Craig (1979, 1980, 1981).

Active criterion adjustment need not be the only process going on during vigilance, and it is important to realize that criterion adjustment may be perfectly compatible with changes in signal discriminability. For example, in a study by Binford and Loeb (1966) of vigilance performance over nine successive sessions, stable probability matching is found during the last two sessions (see Craig, 1978, 1983). (It is data from this Binford study that are depicted in Figure 3.) But as Loeb (1978) has pointed out, discriminability declined within these sessions, implying that the criterion was being adjusted in a way that maintained probability matching. A similar effect has been explicitly obtained by Williams (1986), who trained observers for several sessions until their response rate stabilized and met the criterion of probability matching. During subsequent testing the response rate remained stable at the matching

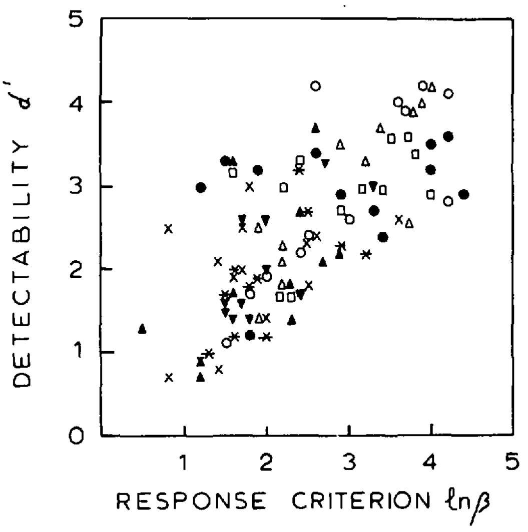

Figure 4. *The association between d' and β in vigilance for a low-probability event. Each point represents the data of a single observer within a group; the different symbols correspond to different groups whose data have been linearly adjusted to remove between-group variation in mean scores.*

level but discriminability was exactly offset by a relaxing of the response criterion, so that responses continued to be made at the previously established rate. Leary (1983) reports an interesting counterexample in which it appears that the rate of responding is adjusted as a reaction to experimenter-controlled changes in discriminability, as if to maintain the likelihood-ratio criterion at a stable level.

CONCLUSIONS

These data signify that attention can be sustained—if it were not sustained, the appropriate adjustments to report rate could not be made. To describe vigilance performance in terms of SDT measures seems sound, and this implies that interpretations derived from SDT applications are probably valid.

In considering probability matching I have stressed the quantitative properties of report

Downloaded from hfs.sagepub.com at PENNSYLVANIA STATE UNIV on September 11, 2016

rate. But what should be underscored is that the observer is an active participant, not a passive bystander, who controls his or her behavior within the degrees of freedom that remain available. In the present context the observer's tendency to limit use of the response rating scale and to adjust report rate up or down as appropriate reflects this control of one's own behavior.

REFERENCES

Baker, C. H. (1963). Further toward a theory of vigilance. In D. N. Buckner and J. J. McGrath (Eds.), *Vigilance: A symposium* (pp. 137–154). New York: McGraw-Hill.

Binford, J. R., and Loeb, M. (1966). Changes within and over repeated sessions in criterion and effective sensitivity in an auditory vigilance task. *Journal of Experimental Psychology, 72*, 329–345.

Broadbent, D. E. (1971). *Decision and stress*. London: Academic Press.

Colquhoun, W. P., and Baddeley, A. D. (1967). Influence of signal probability during pre-training on vigilance decrement. *Journal of Experimental Psychology, 73*, 153–155.

Craig, A. (1977). Broadbent and Gregory revisited: Vigilance and statistical decision. *Human Factors, 19*, 25–36.

Craig, A. (1978). Is the vigilance decrement simply a response adjustment toward probability matching? *Human Factors, 20*, 441–446.

Craig, A. (1979). Vigilance for two kinds of signal with unequal probabilities of occurrence. *Human Factors, 21*, 647–653.

Craig, A. (1980). Effect of prior knowledge of signal probabilities on vigilance performance at a two-signal task. *Human Factors, 22*, 361–371.

Craig, A. (1981). Monitoring for one kind of signal in the presence of another: The effects of signal mix on detectability. *Human Factors, 23*, 191–197.

Craig, A. (1983). Vigilance and inspection. *Studia Psychologica, 25*, 259–270.

Craig, A., and Colquhoun, W. P. (1977). Vigilance effects in complex inspection. In R. R. Mackie (Ed.), *Vigilance: Theory, operational performance, and physiological correlates* (pp. 239–262). New York: Plenum.

Davies, D. R., and Parasuraman, R. (1982). *The psychology of vigilance*. London: Academic Press.

Fisher, R. A. (1970). *Statistical methods for research workers* (14th ed.). Edinburgh: Oliver and Boyd.

Fisk, A. D., and Schneider, W. (1981). Control and automatic processing during tasks requiring sustained vigilance. *Human Factors, 23*, 737–750.

Green, D. M., and Luce, R. D. (1973). Speed-accuracy trade-off in auditory detection. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 547–569). New York: Academic Press.

Grey, D. R., and Morgan, B. J. T. (1972). Some aspects of ROC curve-fitting: Normal and logistic models. *Journal of Mathematical Psychology, 9*, 128–139.

Jerison, H. J., Pickett, R. M., and Stenson, H. H. (1965). The elicited observing rate and decision processes in vigilance. *Human Factors, 7*, 107–128.

Leary, J. (1983). *Adaptation of decision criteria in vigilance tasks*. Unpublished doctoral dissertation, University of Adelaide, Adelaide, Australia.

Loeb, M. (1978). On the analysis and interpretation of vigilance: Some remarks on two recent articles by Craig. *Human Factors, 20*, 447–451.

Long, G. M., and Waag, W. L. (1981). Limitations on the practical applicability of d' and β measures. *Human Factors, 23*, 285–290.

Luce, R. D., and Green, D. M. (1974). Detection, discrimination and recognition. In E. C. Carterette and M. P. Friedman (Eds.), *Handbook of perception*, vol. II (pp. 299–342). New York: Academic Press.

Nuechterlein, K. H., Parasuraman, R., and Jiang, Q. (1983). Visual sustained attention: Image degradation produces rapid sensitivity decrement over time. *Science, 220*, 327–329.

Parasuraman, R. (1979). Memory load and event rate control sensitivity decrements in sustained attention. *Science, 205*, 924–927.

Swets, J. A., and Pickett, R. M. (1982). *Evaluation of diagnostic systems: Methods from signal detection theory*. New York: Academic Press.

Taylor, M. M. (1967). Detectability theory and the interpretation of vigilance data. In A. F. Sanders (Ed.), *Attention and performance I* (pp. 390–399). Amsterdam: Elsevier Science Publishers.

Thorndike, E. L. (1926). *Educational psychology, vol. III. Mental work and fatigue, and individual differences and their causes*. New York: Teachers College, Columbia University.

Vickers, D. (1979). *Decision processes in visual perception*. London: Academic Press.

Vickers, D., and Leary, J. N. (1983). Criterion control in signal detection. *Human Factors, 25*, 283–296.

Vickers, D., Leary, J., and Barnes, P. (1977). Adaptation to decreasing signal probability. In R. R. Mackie (Ed.), *Vigilance: Theory, operational performance and physiological correlates* (pp. 679–703). New York: Plenum.

Williams, P. S. (1986). Processing demands, training, and the vigilance decrement. *Human Factors, 28*, 567–569.

Williges, R. C. (1976). The vigilance increment: An ideal observer hypothesis. In T. B. Sheridan and G. Johannsen (Eds.), *Monitoring behavior and supervisory control* (pp. 181–190). New York: Plenum.

Downloaded from hfs.sagepub.com at PENNSYLVANIA STATE UNIV on September 11, 2016