A Distributed Signal Detection Theory Model: Implications for the Design of Warnings

Jason D. Papastavrou
Mark R. Lehto
Purdue University

A distributed signal detection theory model is employed to analyze the effectiveness of warnings under different operating conditions. In particular, the following two cases are examined: (a) the warning on a product is always present and (b) the warning on a product is administered selectively. The comparative effects of warning versus no warning are described. It is established that selectivity always increases effectiveness. The implications to optimal warning design of intermittent hazard versus continuous hazard are discussed. Furthermore, a series of experiments is conducted to compare the behavior of human participants with the prescriptive behavior of the normative model. The changes in the behavior of the human participants response to changes in the warning levels are consistent with the predictions of the model. These changes should be taken into consideration in the design of warnings.

1. INTRODUCTION AND MOTIVATION

The safe and effective use of many products is predominantly determined by the decisions and actions people make. In general, these decisions and actions are appropriate, but occasionally errors occur that result in serious accidents. Warnings provide information intended to reduce such errors and are often explicitly designed into products.

From an engineering viewpoint, the primary reason to provide a warning is to reduce accidents. The primary assumption underlying this approach is that people will both notice and heed warnings. The majority of warning research and subsequently developed guidelines have been focused on perceptual issues. Decision making and other cognitive aspects of human behavior have been neglected (Lehto & Papastavrou, 1993). Consequently, just as safety communication campaigns in the 1960s were often assumed without foundation to be effective (Haskins, 1969, 1970), warnings are often alleged to be effective or ineffective on the basis of little or no research (Lehto & Miller, 1986).

Although the available literature is limited, there is evidence that warnings should be selective (Lehto & Miller, 1986). Part of the reason is that people often seem to ignore warnings. For example, in a classic experiment, none of 100 participants noticed the explicit warning labels placed on hammers, one of which was the warning label supplied by the manufacturer (Dorris & Purswell, 1977). In a study of 52 participants using 60 different consumer products, it was found that, 34% of the time, participants stated they would not read any of the instructions that came with a product; 53% of the time the participants said they would read all the instructions (Wright, Creighton, & Threlfall, 1982). The average time of warning information in tobacco advertisements viewed by adolescents has been shown to be only 8% of the total viewing time, and in 43.6% of the cases the warning was not viewed at all (Fischer, Richards, Berman, & Krugman,
1989). It has also been found that only 42% of women surveyed noticed a tampon warning when changing from one brand to another (Godfrey & Laughery, 1984).

Several additional experiments indicate that even traffic signs are frequently filtered. In one study, it was found that 15% to 30% of motorists did not recall seeing forest fire safety signs (Ruchel & Folkman, 1965). The average recall by motorists of the last two road signs they passed (they were stopped 200 m away from the signs) has been as low as 4.5% and 16.5% during the day and night, respectively (Shinar & Drory, 1983). Other research showed sign recall levels varying from 21% to 79% for motorists stopped 710 m after passing a traffic sign (Johansson & Backlund, 1970). The studies of traffic signs provide the clearest evidence of filtering because in a related study motorists failed to notice only 2.95% of the passed signs when they were explicitly asked to look and then report the signs to an investigator in the back seat of the car (Summula & Näätänen, 1974).

These studies do not, however, determine when warnings will attract attention rather than be ignored. Many factors can theoretically influence this process. Among such factors are the perceived risk and importance of the warning. One of the most salient findings is that people are more likely to take safety precautions if they believe the danger is large. This principle was confirmed in an experiment in which participants were more likely to behave in accordance with a warning label on a power saw when it was perceived as being highly dangerous (Otsubo, 1988). One review of the research regarding the response of people to volcano, flood, and nuclear power plant warnings concluded that people’s belief that the danger was real (i.e., after officials or police provided warnings) was a very major determinant of behavior. If people did not believe the warning (i.e., when newspapers reported problems), they were much less likely to behave in accordance with the warning (Perry, 1983). Other researchers have emphasized that credible sources are more likely to be persuasive (McGuire, 1980). Similar effects have been shown for public utility customers in which messages stated to come from the public service commission were more likely to elicit effects than those from the electrical utility (Craig & McCann, 1978).

From a decision-making perspective, warnings may also be ignored if they seem to be irrelevant to task performance. Experienced workers in particular may be more prone to ignore warning-related information because of past, benign experience in which accidents rarely occur. Along these lines, it has been hypothesized that people may rationally ignore safety-related advice when the probability of an accident is perceived to be low (Slovic, Fischhoff, & Lichtenstein, 1978). Experimental support for both theories is given by results in which the tendency to read instructions increased when people were unfamiliar with a product or when a product was perceived to be complex, unsafe, or expensive. Complexity and frequency of use correlated significantly with the propensity to read instructions ($r = 0.47$ for the former and $r = -20.24$ for the latter; Wright et al., 1982). Similar results have been obtained for the reading of warning labels (Otsubo, 1988). Drivers have also been found to most likely recall seeing traffic signs perceived as being important (Johansson & Backlund, 1970).

Another concern is that the excessive provision of warnings may result in information overload. There is evidence that increasing the number of items on a label or sign can cause a division of processing time among the items presented. For example, it has been found that participants remember important product-related information better when fewer items are listed on labels (Scammon, 1977). Other research has shown that false alarms have been associated with negative perceptions of pilots toward ground fault warning systems (Loomis & Porter, 1982). A complete overview of the field can be found in Lehto and Papastavrou (1993).

On the basis of such research findings, it has been concluded that there is a need to selectively provide critical information in all forms of warnings, including warning signs or labels (Lehto & Miller, 1986). However, the experimental evidence is not as useful as may be desired. One major shortcoming is that the experimental data does not define the concept of selectivity precisely, much less how selective warnings should be. The experimental evidence may also be misinterpretable as implying that warnings are never of value. This article is intended to go beyond the experimental evidence and show that the need for selectivity is based on fundamental theoretical grounds. Most importantly, it is oriented toward the ultimate development of an objective design methodology in which the optimum level of selectivity is
precisely defined. In attempting to attain this goal, it provides a new and original perspective on the role of the designer.

In this new perspective, the designer becomes a consultant who provides information through an intermediary (the product warning) to the user. Once the product is in use, the product warning becomes in effect the consultant. From the modeling perspective, the warning and user become decision makers who jointly try to make an optimal decision. The decision, of course, is one of determining whether danger is or is not actually present in a situation in which both failures to identify a hazard and false alarms have associated costs. The critical question to the designer is to determine how selective the warning should be to minimize the expected cost to the user as a function of (a) the false alarms and correct identifications made by the warning and (b) the marginal probability of the danger being present. The selectivity issue, when posed in this way, falls within the problem of optimal distributed decision making, as discussed later.

In the following section, the signal detection theory model for the warning process is defined, explained, and analyzed. In Section 3, the experiments are described and the results are discussed. In Section 4, the conclusions are presented.

2. A HYPOTHESIS TESTING PARADIGM

2.1. The Model
The problem of distributed decision making in a hypothesis testing environment has attracted considerable interest during the past decade. This framework was selected because it combines two desirable attributes: (a) The mathematical problems are easy to describe so that researchers from diverse disciplines can understand the models and their conclusions and (b) the problems have trivial centralized counterparts so that all the difficulties arise because of the decentralization of the decision-making process. On the other hand, these problems are also known to become computationally intractable (NP-hard) even for a small number of decision makers (DMs) and a small number of communication messages (Tsitsiklis & Athans, 1985). A thorough overview of the field was presented in Tsitsiklis (1993).

The distributed binary hypothesis testing model for our problem is defined as follows. There are two hypotheses, \( H_0 \) and \( H_1 \), with known prior probabilities, \( P(H_0) > 0 \) and \( P(H_1) > 0 \), respectively, and the team (organization) consists of two DMs: One is called the primary DM and the other is called the consultant DM (Figure 1). Let \( y_n \), the observation of the \( n \)th DM, be a random variable taking values in a set \( Y_n \). It is assumed that the \( y_n \)'s are conditionally independent given either hypothesis, with a known conditional distribution \( P(y_n | H_j), j = 0, 1 \).

The final decision \( u_p \) (0 or 1, indicating \( H_0 \) or \( H_1 \) to be true) is the responsibility of the primary DM. The consultant DM evaluates a binary message \( u_c \) (0 or 1) as a function of its own observation; that is, \( u_c = \gamma_c(y_c) \) in which the measurable function \( \gamma_c : Y_c \to \{0, 1\} \) is the decision rule of the consultant DM. The decision \( u_p \) (0 or 1) of the primary DM is the final team decision and declares one of the hypotheses to be true; that is, \( u_p = \gamma_p(y_p) \) in which again the measurable function \( \gamma_p : Y_p \to \{0, 1\} \) is the decision rule of the primary DM. The objective of the team is to choose the decision rules \( \gamma_n \) for the two DMs, which minimize the probability of error of the team decision, taking into account different costs for hypothesis misclassifications. Let \( J(u, H_i) \) be the cost of the team deciding \( u \) when the true hypothesis is \( H_i \) \((i = 0, 1)\) and define the decision threshold \( \eta \) as follows:

\[
\eta = \frac{P(H_0)[J(1, H_0) - J(0, H_0)]}{P(H_1)[J(0, H_1) - J(1, H_1)]}
\]

1Throughout this discussion it is assumed that the cost function is such that it is more costly for the team to err than to be correct (i.e., \( J(0, H_1) > J(1, H_1) \) and \( J(1, H_0) > J(0, H_0) \)). This logical assumption is made in order to express the optimal decision rules in the convenient likelihood ratio form.
As \( \eta \) varies from zero to infinity, a curve describing the probability of detection (i.e., the probability of deciding \( u = 1 \) when \( H_1 \) is true) versus the probability of false alarm (i.e., the probability of deciding \( u = 1 \) when \( H_0 \) is true) is obtained parametrically. This is called the receiver operating characteristic (ROC) curve and is the cornerstone of mathematical binary hypothesis testing because it offers a complete description of the DMs (Van Trees, 1968). Two important properties of ROC curves that will be explored in the subsequent analysis follow:

1. The better the DM, the higher the DM's ROC curve. This is because, for a given level of false alarm, the DM with the higher ROC curve can achieve a higher probability of detection.

2. The slope of the tangent (if it exists) to a ROC curve at a point \((PF, PD)\) is equal to the value of the threshold \( \eta \) that corresponds to that point. At a point \((PF, PD)\) in which a ROC curve is not differentiable, the value of \( \eta \) that corresponds to that point is a subgradient of the ROC curve at \((PF, PD)\).

**Proposition 1.** The optimal decision rules for the DMs of the tandem team of Figure 1 have to satisfy the following necessary conditions. For the primary DM:

\[
\text{If } u_c = 0: \quad A(y_p) \geq \frac{u_p = 1}{u_p = 0} \quad \frac{1 - P_F}{1 - P_D} \eta = \eta_0.
\]

\[
\text{If } u_c = 1: \quad A(y_p) \geq \frac{u_p = 1}{u_p = 0} \quad \frac{P_F}{P_D} \eta = \eta_1.
\]

For the consultant DM:

\[
A(y_c) \geq \frac{u_c = 1}{u_c = 0} \quad \frac{P_D^1 - P_D^0}{P_D^1 - P_D^0} \eta = \eta_c
\]

in which

\[
A(y_n) = \frac{P(y_n | H_1)}{P(y_n | H_0)}
\]

is the likelihood ratio of DM \( n \) \((n = p, c)\), and in which \( \eta \) was defined in Equation 1 with \( P_D^1 \) and \( P_D^0 \), respectively, the probability of detection and probability of false alarm for the primary DM when \( u_c = i \) was sent by the consultant DM \((i = 0, 1)\) and \( P_F^1 \) and \( P_F^0 \), respectively, the
probability of detection and probability of false alarm for the consultant DM. Moreover, the notation:

\[ u = 1 \]
\[ A(y) \geq \eta \]
\[ u = 0 \]

implies that if \( A(y) > \eta \) the optimal decision is \( u = 1 \), if \( A(y) < \eta \) the optimal decision is \( u = 0 \), and if \( A(y) = \eta \) the optimal decision is either \( u = 0 \) or \( u = 1 \).

The analytical proof can be found in Papastavrou and Athans (1992).

**Remark 1.** The optimal decision thresholds of the two DMs are given by a set of coupled equations. For example:

\[ P' = P \left( \frac{1 - P_F}{1 - P_D} \eta \right| H_0) \]  

This implies that changes in the decision rule of one DM induce changes in the decision rule of the other DM: the optimal decision rule of each DM not only on the given parameters that are external to the team (i.e., costs and prior probabilities) but also on who the DMs decision-making partner is. Thus, we cannot refer to the optimal decision rule of a (single) DM but rather of the set of optimal decision rules of the DMs in the team.

**Remark 2.** Let \( P'_D(\eta) \) and \( P'_F(\eta) \), respectively, denote the probability of detection and probability of false alarm for the team as a whole for some given value of the threshold \( \eta \). The ROC curve of the team as a whole can be computed and is given by the following two parametric equations:

\[ P'_D(\eta) = [1 - P'_F(\eta)]P'_D(\eta) + P'_F(\eta) P'_D(\eta) \]  

\[ P'_F(\eta) = [1 - P'_D(\eta)]P'_D(\eta) + P'_D(\eta) P'_F(\eta) \]  

Note that the team ROC curve depends not only on the characteristics (expertise) of the individual DMs, but also on the particular way that they have been constrained to interact (the team or organization architecture). That is, the team ROC curve depends on both the particular form of the ROC curves of the individual DMs and on which DM has been designated to be the primary DM.

### 2.2. The Consultant DM as a Warning Message

Equations 2 and 3 give the optimal actions of the primary DM, given that the consultant DM employs \( (P_F, P_D) \) as its operating point; that is, Equations 2 and 3 indicate the optimal way in which the primary DM can benefit from the consultant’s communication. This does not depend on whether the operating point is optimal or suboptimal from the consultant DM’s point of view, or, for that matter, on whether \((P_F, P_D)\) was selected according to some arbitrary rule.

Therefore, the team of Figure 1 can also be viewed in a different context: The consultant DM provides a warning to the primary DM about some imminent danger (represented by hypothesis \( H_1 \)). The team still consists of cooperating DMs but now the consultant DM does not strive to optimize its decision; rather, some external conditions to the team may dictate the operating point of the consultant DM (e.g., a law that requires that a warning label or sign always be present). The primary DM still optimally combines its personal observation with the communication from the consultant DM so as to optimize the objective function (i.e., the expected cost of the team decision). Therefore, only Equations 2 and 3 will be necessarily satisfied.
Suppose that the decision threshold is given to be \( \eta^* \). Consider the effects of the movement of the consultant DM's operating point along the ROC curve of the consultant DM (Figure 2). Suppose that the operating point of the consultant DM is at (0, 0); this implies that, independent of its observation \( y_c \), the consultant DM is always sending \( u_c = 0 \) to the primary DM. Then, the optimal operating point of the primary DM \((P_D^0, P_F^0)\) is at \((P_D^0, P_F^0)\) in which \((P_D^0, P_F^0)\) is the maximum likelihood operating point of the primary DM if it was deciding in isolation for the same value of \( \eta^* \) and the other operating point \((P_D^1, P_F^1)\) of the primary DM is at (1, 1, but this is never used because \( u_c = 1 \) is never communicated by the consultant DM. That is, in this case, the primary DM completely disregards the message from the consultant DM and makes the exact same decision that it would be making if it operated alone optimally.

Similarly, suppose that the operating point of the consultant DM is at (1, 1). This implies that, independent of its observation \( y_c \), the consultant DM is always sending \( u_c = 1 \) to the primary DM. Then, the optimal operating point \((P_D^1, P_F^1)\) of the primary DM is at \((P_F^1, P_D^1)\) in which \((P_F^1, P_D^1)\) is the maximum likelihood operating point of the primary DM if it was deciding in isolation for the same value of \( \eta^* \) and the other operating point \((P_D^0, P_F^0)\) of the primary DM is at (0, 0, but this is never used because \( u_c = 0 \) is never communicated by the consultant DM. That is, in this case as well, the primary DM completely disregards the message from the consultant DM and makes the exact same decision that it would be making if it operated alone optimally.

The optimality condition for the consultant DM's operating point is:

\[
m_c = \frac{P_F^1 - P_D^0}{P_D^1 - P_F^0} \eta^*
\]

Figure 2. Effects of movement of the consultant DM's operating point along the ROC curve.
in which $m_c$ is the slope of the tangent at the consultant DM's operating point. As the operating point moves along the consultant DM's ROC curve, the slope of its tangent decreases from infinity to zero in a continuous manner. At the same time, the right part of Equation 9 also decreases continuously from a finite positive value to another (smaller) finite positive value. Thus, the existence of optimal decision rules can also be established this way. Moreover, the fact that the necessary optimality conditions are not sufficient can be obtained because Equation 9 can have more than one solution.

2.3. The Optimal Team Operating Point: A Geometric Approach

Suppose that the threshold is given to be $\eta^*$ and that the optimal operating point of the consultant DM, $(P_F^c, P_D^c)$, has been determined; for example, the ROC curve of a smoke alarm system is specified during development. The ROC curve of the consultant DM is presented in Figure 3. Then, according to Equations 2 and 3, the optimal operating points $(P_F^p, P_D^p)$ and $(P_F^t, P_D^t)$ of the primary DM are also obtained. Equations 7 and 8 can be used to obtain the operating point of the team.

There exists an alternative geometric way to determine the operating point of the team. This is demonstrated in Figure 4. First, consider the ROC of the consultant DM (Figure 3) and scale the false alarm axis by multiplying with $(P_F^c - P_D^c)$ and scale the detection axis by multiplying with $(P_F^c - P_D^c)$. Then, place this scaled version of the consultant DM's ROC curve on top of the primary DM's ROC curve so that its southwest point of the scaled version (originally the point $(0, 0)$) is on $(P_F^c, P_D^c)$ of its northeast point (originally the point $(1, 1)$) is on $(P_F^c, P_D^c)$. The location of $(P_F^c, P_D^c)$, the consultant DM's operating point, in this scaled version is $(P_F^t, P_D^t)$, the optimal operating point for the team. By construction, $(P_F^t, P_D^t)$ always lies above the ROC curve of the primary DM and thus the following observations can be made:

- The team can achieve a performance that neither DM can achieve in isolation.
- The primary DM's performance is not improved because of the presence of the message from the consultant DM. Moreover, the primary DM does not have two ROC curves, one when $u_c = 0$ and one when $u_c = 1$ is received from the consultant DM. The primary DM, instead, operates at two points, $(P_F^p, P_D^p)$ and $(P_F^t, P_D^t)$, that belong to the primary DM's (original) ROC curve. Importantly, the consultant DM uses the message transmitted to the primary DM to tell the primary DM which of the two operating points to employ; thus, the overall team performance is superior to the performance of each DM in isolation.
Now consider another value of the threshold, say $\eta^*$, and suppose again that $(P_F^*, P_D^*)$, the corresponding optimal operating point for the consultant DM, has been determined as well. Without loss of generality assume that $P_F^* \leq P_F^i$ and $P_D^* \leq P_D^i$ (Figure 3). This in turn implies, from Equations 2 and 3, that $P_F^i \leq P_F^j$ and $P_D^i \leq P_D^j$, and $P_F^i \leq P_F^j$ and $P_D^i \leq P_D^j$ in which $(P_F^i, P_D^i)$ is the optimal operating point of the primary DM when the consultant DM transmits the message $u_c = i$ ($i = 0, 1$) in this case. All these operating points have been added to Figure 4 and are presented in Figure 5. Then $(P_F^*, P_D^*)$, the optimal operating point of the team, can be obtained using the geometric approach described earlier. From the previous analysis, the following conclusion can be drawn:

- The team ROC curve can be determined by connecting all the points $(P_F^i, P_D^i)$ that are obtained using the geometric method for all the values of the threshold $\eta = 0$.

2.4. The Team Performance Versus the Performance of the Warning Message

Consider again the case in which the consultant DM is viewed as a warning sign. The warning sign is addressed to all the potential users of a product and does not change from user to user. This implies that the consultant DM sends a general warning message and does not optimize for each and every primary DM. The consultant DM should decide on an operating point that would maximize the total “society” objective function. Thus, the designer of the warning should choose an operating point that seems reasonably close to being optimal. On the other hand, given this operating point, each and every user should strive to maximize their personal benefit from this warning; this implies that, given the operating point $(P_F^i, P_D^i)$ of the consultant DM, each user has to operate at the points described by Equations 3 and 4. It would, therefore, be interesting to examine how the team probabilities of false alarm and of detection change as the operating point of the warning sign changes from $(0, 0)$ to $(1, 1)$.

![Figure 4. The geometric approach for obtaining the optimal team operating point.](image-url)
The team probability of false alarm, which is given by Equation 7, is considered first. As was already discussed in the previous section, assuming that the threshold $\eta$ remains constant throughout, at both boundary points of the consultant DM (viz., (0, 0) and (1, 1)), the probability of false alarm of the team is $P_F$ in which $P_F$ is the maximum likelihood probability of false alarm of the primary DM for the given $\eta$. It would be interesting to determine whether $P_F$, the team probability of false alarm, changes in a concave way with $P_p$, the probability of false alarm of the consulting DM. For this, it has to be determined whether the second partial derivative of $P_F$ with respect to $P_p$ is negative. Then:

$$\frac{\partial^2 P_F}{\partial P_p^2} = \frac{\partial^2 \left( (1 - P_F) P_p^0 + P_F P_p^1 \right)}{\partial (P_F)^2} = \frac{\partial^2 P_F}{\partial P_p^2} + 2 \frac{(\partial P_F^1 - P_F^0)}{\partial (P_F)^2} + P_F \frac{\partial^2 P_F}{\partial (P_F)^2}$$

Both the first partial derivatives in the equations are negative. It will only be shown that the partial derivative of $P_F$ with respect to $P_p$ is negative because the proof for the other is very similar. Because of the concavity of the consultant's ROC curve, as $P_F$ increases the ratio:

$$\frac{1 - P_D}{1 - P_F}$$

decreases. Consequently, the threshold $\eta_0$ of the primary DM (given by Equation 2) increases. Because of the property of the tangent to the ROC curve and of the concavity of the ROC curve, it is obtained that $P_F$ decreases which in turn implies that the partial derivative of $P_F$
with respect to $P_F$ is indeed negative. Still, the signs of the second derivatives cannot be determined as they depend on the particular shape of the ROC curves of the DMs. Thus, $P_F$ cannot be shown to be concave with respect to $P_F$. In fact, examples exist to show that it is not concave.

This result implies that, as the probability of false alarm of the warning message varies, there are no a priori mathematical inferences that can be made about the change of the probability of false alarm of the primary DM; the probability of false alarm of the primary DM does not change in a nice and predictable pattern. A similar result can be derived for the probability of detection of the primary DM.

2.5. Discussion

2.5.1. Inefficient Warning Mechanisms

It is clear from the discussion of Section 2.2 that, if the primary DM behaves optimally, a warning message that does not contain any information about the true state of the environment does not affect the primary DM’s behavior; this includes the permanent presence (or absence) of a warning sign. Consequently, the system that consists of the warning message and the primary DM performs suboptimally because the warning message is effectively ignored. There are two different ways to realize this:

1. The presence of the consultant DM (i.e., warning mechanism) does not improve the quality of the decision of the decision-making team. The primary DM will always be operating at the point $(P_F, P_D)$ of its ROC curve, which is the optimal point which the primary DM would employ if it had to make a decision alone. Therefore, the team operating point is also the point $(P_F^*, P_D^*)$, which can be achieved by the primary DM working in isolation.

2. It can also be seen mathematically because at the points $(P_F, P_D) = (0, 0)$ and $(P_F, P_D) = (1, 1)$, the necessary optimally condition Equation 9 is not satisfied.

Therefore, even if the warning mechanism cannot or does not optimize the performance of its decision, warnings should be administered selectively in order to be effective; otherwise, assuming that DMs are optimizers, warnings will not have any effect on the decision of the primary DM as they will be ignored.

2.5.2. The Infallible Warning Mechanism

Suppose that the warning mechanism can always perform perfect detection; that is, suppose that $(P_F, P_D) = (0, 1)$. Then, the optimal decision rule for the primary DM is to always obey the warning message (and thus ignore its own observation). In fact, from Equation 2 it can be obtained that $(P_F^*, P_D^*) = (0, 0)$, which implies that whenever the primary DM receives a message $u_c = 0$, it also decides $u_p = 0$. Similarly, from Equation 3 it can be obtained that $(P_F^*, P_D^*) = (1, 1)$, which implies that whenever the primary DM receives a message $u_c = 1$, it also decides $u_p = 1$. Therefore, the team always achieves perfect detection.

2.5.3. The Primary DM Without a Personal Observation

Suppose that the primary DM does not have a personal observation of the environment or that its observation is very bad; in that case, its ROC curve is the diagonal from $(0, 0)$ to $(1, 1)$. In this case, the optimal performance that the team can achieve is the best performance that the warning message can achieve if it operated in isolation. It can be shown that, as in the previous case, it will be optimal for the primary DM to always accept the warning message as the team’s decision.

2.5.4. Optimizing the Performance of the Team

The final result should be especially important in the design of warnings because it seems logical that, in order to increase the probability of detection of the primary DM, the probability of detection of the warning (i.e., the consultant DM) has to increase. But it was shown that, as the probability of detection of the warning increases, it does not necessarily cause the probability of detection of the warning or primary team to increase as well. Similarly, in order to decrease the probability of false alarm of the primary DM, the probability of false alarm of the
warning (i.e., the consultant DM) has to decrease; again, it was shown that, as the probability of false alarm of the warning increases, it does not necessarily cause the probability of false alarm of the warning or primary team to increase as well. Therefore, in order to modify the level of the performance (i.e., the operating point) of some primary DM in a desired way, the designer of the warning needs to take into consideration both the particular characteristics (i.e., the ROC curve) and the current level of performance of the primary DM.

3. THE EXPERIMENT

An experiment was performed to evaluate the degree to which the predictions of the model hold true for human participants. The focus was on examining whether changes in the warning threshold affected the participants’ behavior as predicted by the distributed signal detection theory model; that is, whether the participants’ operating points shifted as predicted in Section 2.2. More specifically, as the operating point of the warning moves from (0, 0) to (1, 1), a corresponding shift should be seen on the ROC curve of the primary DM. That is, \((P_p^*, P_{g}^*)\) should shift from \((P_p^0, P_{g}^0)\) to \((0, 0)\) and \((P_p^1, P_{g}^1)\) from \((1, 1)\) to \((P_p^*, P_{g}^*)\). In other words, as the warning threshold is reduced, resulting in more frequent warnings, the warning should become less believable.

3.1. Method and Procedures

3.1.1. Participants
Six participants ranging in age from 19 to 31 years old drawn from the population of the School of Industrial Engineering at Purdue University participated in this experiment; each was paid at a rate of $4.50 per hour.

3.1.2. Experimental Task
A user-friendly interactive dual task environment, programmed in HYPERCARD and implemented in a Macintosh computer, was developed. The dual task environment included signal detection as the primary task and digit identification as the secondary task. Both tasks were presented to participants within a single 5 in. \(\times\) 7 in. window on a 14-in.-high resolution CRT monitor, as illustrated in Figure 6, and were performed concurrently.
The secondary task required participants to identify the contents of a display that contained an integer between 1 and 4. To do this, participants pointed and clicked on the appropriately numbered button immediately below the display (see Figure 6). The display was updated every 1.5 s. Participants were given 1.5 s to respond after the display was updated and either gained or lost 10 points depending on whether their response was correct or incorrect. A failure to respond within the time window of 1.5 s was interpreted as an error.

For the primary task, participants pointed and clicked on one of two buttons to indicate whether or not they thought the signal event would occur. These buttons, as shown in the lower right corner of Figure 6, were labeled as $T$ for true and $F$ for false. A central bar-type display provided an imperfect estimate of the probability of the signal event; that is, the height of the bar was equal to the sum of the true probability ($p$) that a signal event would occur and an uniformly distributed noise term ($n$), subject to the constraint that sums below 0 or above 1 were truncated at 0 and 1, respectively. The range of the noise term, $n$, was from $-0.2$ to 0.2.

The interarrival times for signal events have a weighted negative exponential distribution; this consists of a constant wait of 2.25 s plus an exponentially distributed wait with a mean of 1.5 s. At each of these times, the central display provided an updated estimate of the probability of the signal event. A failure to respond within a 5-s time window after the central display updated was interpreted as selecting the response $false$. Participant responses were respectively categorized as a correct identification if the participant correctly detected the presence of a target, a miss if the participant did not detect the presence of a target, a correct rejection if the participant correctly detected the absence of a target, and a false alarm if the participant incorrectly decided that a target was present. Payoffs for each outcome (see Table 1) were added to the participant’s score. Scores and outcomes were displayed on the computer screen (see Figure 6) and updated immediately.

The dual task environment also allowed warnings to be provided at the time the central display for the primary task was updated. The warnings provided could be either visual (in the form of a flash on the screen) or auditory (in the form of a beep). Additionally, the frequency of the warning could be controlled by using different thresholds reflecting the likelihood of the signal event. Importantly, the estimate of the likelihood of the signal event used by the computer when deciding whether to provide a warning was independent but identically distributed to the estimate provided to the human participant in the central display. The computer, in the way it selectively provides warnings, therefore exactly corresponds to the role of the consultant DM in the model discussed in Section 2.

Several different threshold values were used. One level was the optimal threshold that the consultant DM would employ if it had to make the decision by itself (see Equation 1). Because the particular characteristics of each participant (i.e., the ROC curve) are not known a priori, this threshold was selected for the warning message because it performs well independent of the consultant DM. The other threshold values were selected in relation to this threshold value as is described in the Section 3.1.4 later.

### 3.1.3. Data Collected

The computer recorded all participant responses and external events throughout the experiment. Such data included the times at which signal events could occur, the true and presented probability of a signal event, whether a warning was provided, and the participants’ responses (i.e., whether the participant decided that a target was present or not).

<table>
<thead>
<tr>
<th>Event</th>
<th>Notation</th>
<th>Payoff/Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct identification</td>
<td>$J(1, H_1)$</td>
<td>25</td>
</tr>
<tr>
<td>Miss</td>
<td>$J(0, H_1)$</td>
<td>$-50$</td>
</tr>
<tr>
<td>Correct rejection</td>
<td>$J(0, H_0)$</td>
<td>10</td>
</tr>
<tr>
<td>False alarm</td>
<td>$J(0, H_0)$</td>
<td>$-10$</td>
</tr>
<tr>
<td>Secondary detection</td>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>
Using this data, the program then calculated the probability of detection (i.e., the proportion of time that the participant correctly identified the presence of a target) with and without a warning being present, the probability of false alarm (i.e., the proportion of time that the participant incorrectly identified the presence of a target) with and without a warning being present, the total payoffs from the detection process, and the total payoffs from the secondary task. This information was stored on a computer spreadsheet and then analyzed, as will be described later.

3.1.4. Experimental Procedure
Participants received extensive training over a period of 6 hr in six different sessions prior to participating in the experiment. During the training sessions, they learned the costs involved with the various combinations of responses and events, developed basic skills, and learned how to interpret the various displays. After completing their training sessions, the participants then performed their experimental tasks in two different sessions. Each session lasted 2 hr. In the first session, the warnings were administered with a flash on the screen and in the second session the warnings were administered with a beep.

In each of the two sessions, four different levels of warnings were tested: always warn, warn if the consultant's estimate of the probability of the signal event was respectively greater than 8.34%, greater than 16.67%, or greater than 58.34%. These warning levels were not selected randomly; rather, 16.67% is the "break even" point in the sense that is equally costly always to decide that a signal is present and always to decide that there is no signal present. Thus, for the cost matrix used in this experiment (see Table 1):

\[ pJ(1, H_1) + (1 - p)J(1, H_0) = pJ(0, H_1) + (1 - p)J(0, H_0) \Rightarrow \\
\]
\[ p = [J(0, H_0) - J(1, H_0)]/[J(1, H_1) - J(0, H_1) + J(0, H_0) - J(1, H_0)] = 0.1667 \]

Also, 8.34% is half way between the threshold of 16.67% and always warning. Similarly, 58.34% is half way between the threshold of 16.67% and never warning. All the other factors, except the participants and the warning level, were treated as fixed. The order of presentation of the different warning levels was counterbalanced across participants.

3.2. Results
As discussed earlier, four warning levels were presented to each participant. For each warning level, two points on the ROC curve of each participant were obtained: the operating point when a warning is not present (i.e., \((P^*_D, P^*_F)\) in the notation of the previous section) and the operating point when a warning is present (i.e., \((P^*_D, P^*_F)\) in the notation in the previous section). The ROC curves are presented in Figures 7 through 12 and are based on the averaged values for the two types of warnings (this was done because, as will be discussed in Section 3.2, the differences from the two types of warnings were not statistically significant). Some of the ROC curves have less than 8 points because the point \((0, 0)\) was employed more than once (i.e., for more than one warning level). Also because all ROC curves go from \((0, 0)\) to \((1, 1)\), the point \((1, 1)\) was added in some of the ROC curves to preserve the scale.

In order to reduce the noise and statistical variability, the moving average of the points of each ROC curve was also calculated. These "smoothed" ROC curves are also presented together with the ROC curves obtained from the raw data. Because the objective of this experimental study is to establish the existence, or lack, of a pattern in the shifts of the operating points of the participants response to the changes in the warning levels, the use of the smoothed ROC curves is justified. If a pattern indeed exists it will be accentuated; if a pattern does not exist and the shifts are random, then the moving average will converge around the optimal operating point that each participant would employ if the participant had to make the decision in isolation (i.e., \((P^*_D, P^*_F)\) in the notation of the previous section).

From observing the experimental ROC curves it appears that the participants’ performance is quite consistent with that prescribed by the normative signal detection model. In particular,
Figure 7. Experimental ROC curve and performance of Participant 1.

Figure 8. Experimental ROC curve and performance of Participant 2.
A SIGNAL DETECTION THEORY MODEL FOR THE DESIGN OF WARNINGS

Figure 9. Experimental and smoothed ROC curve of Participant 3.

Figure 10. Experimental ROC curve and performance of Participant 4.
Figure 11. Experimental ROC curve and performance of Participant 5.

Figure 12. Experimental ROC curve and performance of Participant 6.
observation of the ROC curves quickly reveals a definite pattern in the shifts of the operating points of the participants. To help interpret this trend, recall that in Section 2.2, we showed that as the operating point of the warning moves from $(0, 0)$ to $(1, 1)$, $(P_F, P_D)$ should shift from $(P_F, P_D)$ to $(0,0)$ and $(P_F, P_D)$ from $(1,1)$ to $(P_F, P_D)$. This suggests that as the frequency of the warning increases (i.e., as the warning threshold decreases), the participants should place less importance in the presence of a warning and more importance in the absence of a warning. The results indeed suggest that participants modify their behavior (i.e., shift their decision thresholds) in response to changes in the warning level as predicted by the signal detection model.

A second test of whether the participants behaved consistently with the predictions of the model is related to the extent the obtained data traces out a concave, nondecreasing ROC curve. This follows because the model predicts that changing the warning threshold conditions will cause shifts on the participants' ROC curve. Evidence for or against this conjecture can be obtained by examining whether the piece-wise linear data in Figures 7 through 12 are consistent with what would be expected if they represent points along a single ROC curve. The approach we took was to evaluate how many segments of the piece-wise linear data obtained for our participants satisfy this criteria and then compare that number to the expected number given that there was no relation. That is, for a linear segment to satisfy the properties of the ROC curve it has to have:

- **positive slope**: If the points are selected randomly this occurs with $p = .25$.
- **slope less steep than the slope of the previous segment**: If the points are selected randomly this occurs with $p = .5$ on the average.

A total of 29 out of 47 segments were found to satisfy these fundamental properties of the ROC curves (see Table 2). Now consider that if the operating points are random, consecutive line segments will satisfy the properties of the ROC curve with $p = .125 (= .25 \cdot .5)$. This is a conservative estimate of the probability because as the ROC curve approaches the point $(1, 1)$ and becomes less steep, it becomes progressively very difficult for the concavity property to be maintained. The interesting result is that given that the probability of success of a Bernoulli random variable is .125, the probability of having 29 or more successes in 47 trials is less than $10^{-13}$. Furthermore, even if the second condition on the slope is ignored, meaning a positive slope is the only remaining condition and thus the probability of success is increased to .25, the probability of 29 or more successes in 47 trials is less than .0287. This analysis strongly supports the existence of a pattern in the shifts of the operating points of the participants consistent with that predicted by the normative model.
Further examination of Figures 7 through 12 does reveal certain minor discrepancies between the observed results and the normative predictions. In particular, Participants 1, 2, 3, and 6 have one point on the axis of the probability of false alarm (i.e., \(P_F > 0, P_D = 0\)). However, it does appear that these points, which occurred only for low frequency for warning, are because they are caused by at most one or two incorrect detections (i.e., false alarms) out of more than 20 tries. Thus, the tails of the ROC curves are sensitive.

Implicit to the signal detection model is that each DM performs at the best of his or her ability. This implies that each DM, who is characterized by a single ROC curve, always selects his or her operating points from the points that belong on the ROC curve. (A DM could conceivably choose to operate at a point under the ROC curve; this would correspond to suboptimal performance because the DM could increase his or her probability of detection without increasing his or her probability of false alarm.)

On the other hand, from the human performance point of view, it may be plausible to argue that each DM is characterized by two ROC curves: one ROC curve when no warning has been issued and a second ROC curve when a warning has been issued. The rational behind this is that humans do not always perform at the best of their ability. When they do not receive a warning, they devote a certain mental effort to the task and their performance is described by a corresponding ROC curve. When they do receive a warning, they devote a greater mental effort to the task and thus achieve superior performance that is described by another ROC curve.

The experimental results do not provide a conclusive answer to this issue. The two participants that performed best (Participants 2 and 3) are very clearly characterized by a single ROC curve; the participants perform at the best of their abilities, thus satisfying the assumption of the signal detection model and verifying the model's prediction. However, it also appears that other participants (like Participants 5 and 6) could be characterized by two ROC curves. We believe that if the participants perform at the best of their abilities all of the time, they will be characterized by a single ROC curve. If they do not perform at the best of their abilities all of the time, then a dual ROC curve description may be in order. This issue should be more carefully analyzed in future research.

Also in Figures 7 through 12, the “team” (i.e., warning and participant) performance is presented for each of the four warning levels. In Section 2.3 it was discussed that, as the probability of false alarm of the warning message varies, there are no a priori inferences that can be made about the change of the probability of false alarm of the primary DM; the probability of false alarm of the primary DM does not change in a nice and predictable pattern. The only prediction that can be made is that the team performance should be at the same level if a warning is never issued (i.e., \((P_{WF}, P_{WD})\) = (0, 0)) and if a warning is always issued (i.e., \((P_{WF}, P_{WD})\) = (1, 1)); in these two cases the level of performance of the team is \((P_{p}, P_{d})\), the optimal operating point if the primary DM performed in isolation. A similar result can be derived for the probability of detection of the primary DM. The results confirmed the unpredictability in the movement of the team operating point in response to the changes in the warning level.

One final note is that the best performance of all participants and warning levels was achieved by Participant 3 at a warning threshold of 0.1667. At this warning level, Participant 3's operating points were \((P_{WF}, P_{WD})\) = (0.037, 0.0) and \((P_{WF}, P_{WD})\) = (0.759, 0.988) and resulted in a corresponding team operating point of \((P_{p}, P_{d})\) = (0.518, 0.953).

### 3.2.1. Auditory Versus Visual Warning

As was mentioned in Section 3.1, the participants were provided with two types of warnings: auditory and visual. Statistical analysis was performed to determine whether there were any

---

2With the exception of Participant 4, whose performance exhibited significant discrepancies, including several reversals of the ROC curve, perhaps indicating a lack of understanding of the task.

3In the framework of the signal detection model, this can be interpreted as follows: Whenever a warning is issued, the primary DM receives a more accurate observation or analyzes more carefully his or her observation. The primary DM is thus able to better distinguish between the hypotheses.
TABLE 3. Analysis of Variance Procedure

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F Value</th>
<th>p &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel</td>
<td>1</td>
<td>0.24</td>
<td>.6284</td>
</tr>
<tr>
<td>Level</td>
<td>3</td>
<td>2.43</td>
<td>.0816</td>
</tr>
<tr>
<td>Participant</td>
<td>5</td>
<td>14.77</td>
<td>.0001</td>
</tr>
<tr>
<td>Channel level</td>
<td>3</td>
<td>1.49</td>
<td>.2349</td>
</tr>
</tbody>
</table>

differences in the performance of the participants between the two types of warnings. The results of the analysis of variance are presented in Table 3. The various sources analyzed were channel (auditory and visual), level (the four different levels of warning, that is 0.0, 0.0834, 0.1664, and 0.5834), and participant (the six participants employed).

As can be seen from Table 3, the main effect of channel was not statistically significant, $F(1, 35) = 0.24, p > .63$), so further analysis was not deemed important, especially because this was not the main focus of the research. Warning level did have a significant main effect ($F(3, 35) = 2.43, p > .08$), as did participant, $F(5, 35) = 14.77, p > .0001$.

4. SUMMARY AND CONCLUSIONS

A signal detection model for the warning process was developed. This provides a powerful mathematical tool for the analysis and design of better warnings. It provides great insight on the manner that the presence or absence of a warning signal should be interpreted optimally.

Some preliminary experiments were performed to determine whether human DMs behave as predicted by the model. The results indicate that humans behave quite consistently (but not optimally) with the model’s predictions. It appears that shifts in the behavior of the participants (i.e., the shifts in the operating points) in response to changes in the warning levels are in accordance with the predictions of the model. Furthermore, if the participants perform at the best of their abilities, they are characterized by a single ROC curve independent of whether a warning is issued or not. It is also interesting to note that, even though the participants had an extensive training period, not all of them identified the break-even warning threshold of 0.1667. This is very important to the designers of warnings who should be aware that not all consumers use the warnings the way that are supposed to.

We continue to perform experiments to further examine the effects of warning in human behavior. Also, we intend to determine to what degree these results, that are obtained in dynamic computer simulated environment, can be generalized and, therefore, be valid in a real-life environment. In order to achieve this, we are ready to conduct experiments with a truck driving simulator. The main task will be for the participants to safely pass slower moving vehicles; a warning will be provided to assist the participants in doing so. Our ultimate objective is to develop the best procedure for using warning to influence human behavior toward the optimal behavior as predicted by the signal detection model.

REFERENCES


