MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence
Memo. No. 175

April 1969

# On Optimum Recognition Error and Reject Tradeoff*

C. K. Chow

The performance of a pattern recognition system is characterized by its error and reject tradeoff. This paper describes an optimum rejection rule and presents a general relation between the error and reject probabilities and some simple properties of the tradeoff in the optimum recognition system. The error rate can be directly evaluated from the reject function. Some practical implications of the results are discussed. Examples in normal distributions and uniform distributions are given.

*Technical Report by
IBM Watson Research Center
Yorktown Heights, New York

## Introduction

The error rate and the reject rate are commonly used to describe the performance level of pattern recognition systems. An error of misrecognition occurs when a pattern from one class is identified as that of a different class. The error is sometimes referred to as a substitution error or undetected error. A reject occurs when the recognition system withholds its recognition decision, and the pattern is rejected for exceptional handling, such as rescan or manual inspection.

Because of uncertainties and noise inherent in any pattern recognition task, errors are generally unavoidable. The option to reject is introduced to safeguard against excessive misrecognition; it converts potential misrecognition into rejection. However, the tradeoff between the errors and rejects is seldom one for one. Whenever the reject option is exercised, some would-be correct recognitions are also converted into rejects. We are interested in the best error-reject tradeoff in the optimum rejection scheme.

An optimum rejection scheme was derived in Ref. 1. The error-reject tradeoff curves have been used to describe and compare the empirical performances of recognition methods.

(e.g. Refs. 2 and 3), and they have also been found useful in the actual system design of an optical page reader (Ref. 4). However, few theoretical results on the error-reject tradeoff are available.

This paper first describes an optimum rejection rule and then derives a general relation between the error and reject probabilities. The error rate can be directly evaluated from the reject function. This result provides a basis for calculating the error rates from the empirical rejection curve without actually identifying the errors. Some simple properties of the optimum tradeoff are presented. Examples in normal distributions and uniform distributions are given.

where v is the pattern vector, n is the number of classes, $(p_1, p_2, \ldots p_n)$ is the a priori probability distribution of the classes, $F(v|i)$ is the conditional probability density for v given the $i^{th}$ class, $d_i (i \neq 0)$ is the decision that v is identified as of the $i^{th}$ class while $d_o$ is the decision to reject, and t is a constant between 0 and 1 $(0 \leq t \leq 1)$. The probability of error, or error rate, is

$$E(t) = \int_V \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \delta\,(d_j|v) p_i F(v|i) dv \qquad (5)$$

and the probability of reject or reject rate, is

$$R(t) = \int_V \delta(d_o|v) \sum_{i=1}^{n} p_i F(v|i) dv \qquad (6)$$

where V is the pattern space. Both the error and reject rates are implicit functions of the parameter t.

The probability of correct recognition is

$$C(t) = \int_V \sum_{i=1}^{n} \delta\,(d_i|v) p_i F(v|a_i) dv$$

$$= 1 - E(t) - R(t) \qquad (7)$$

and the probability of acceptance (or acceptance rate) is defined as

$$A(t) = C(t) + E(t). \tag{8}$$

Rejection Threshold

The parameter t in the decision rule will be called "the rejection threshold". For any fixed value of t ($0 \leq t \leq 1$) the decision rule $\delta$ partitions the pattern space V into two disjoint sets (or regions) - $V_A(t)$ and $V_R(t)$ where equations (2) and (3) respectively hold, namely:

$$V_A(t) = \{v \mid \max_i [p_i F(v|i)] \geq (1-t)\ F(v)\} \tag{9}$$

$$V_R(t) = \{v \mid \max_i [p_i F(v|i)] < (1-t)\ F(v)\} \tag{10}$$

where

$$F(v) = \sum_{i=1}^{n} p_i F(v|i). \tag{11}$$

Without loss of generality, it will be assumed that F(v) is nonzero over the entire space V, otherwise the set over which F(v) is zero is first deleted. $V_A$ and $V_R$ are called respectively the "acceptance region" and the "reject region" of the decision rule. An example is depicted in Fig. 1(a) where the shaded region is $V_R$ and the unshaded region is $V_A$.

We shall now present some simple properties of the rejection threshold t:

(a) both the error and reject rates are monotonic in t,

(b) t is an upper bound of the error rate, and

(c) t is a differential error-reject tradeoff ratio.

(a) Monotonicity

It follows immediately from the definitions of (9) and (10) that for any $t_1$ and $t_2$ in $[0,1]$ if $t_1 < t_2$, then

$$V_A(t_1) \subset V_A(t_2) \quad \text{and}$$

$$V_R(t_1) \supset V_R(t_2).$$

With the aid of equations (1) and (4), (9) and (10), the various probabilities can be written as

$$R(t) = \int_{V_R(t)} F(v)dv \tag{6'}$$

$$A(t) = \int_{V_A(t)} F(v)dv \tag{8'}$$

and

$$C(t) = \int_{V_A(t)} \max_i [p_i F(v|i)]dv \tag{7'}$$

$$E(t) = \int_{V_A(t)} \{ \sum_{i=1}^{n} p_i F(v|i) - \max_i [p_i F(v|i)] \} dv \tag{5'}$$

All the integrands in the above integrals are non-negative, hence if the domain of integration expands, the value of integral increases. More specifically, if $t_1 < t_2$, then $V_A(t_1) \subset V_A(t_2)$ and $V_R(t_1) \supset V_R(t_2)$; therefore, $E(t_1) \leq E(t_2)$ and $R(t_1) \geq R(t_2)$.

In other words, E increases and R decreases with increasing t. In particular, when $t = 0$, $E = 0$ and when $t = 1$, $A = 1$ and $R = 0$. Whenever $t \geq 1 - \frac{1}{n}$, $R = 0$.

(b) An Upper Bound of Error Rate

We shall now show that

$$E(t) \leq t . \tag{12}$$

For any v in $V_A(t)$, we have

$$\max_i [p_i F(v|i)] \geq (1-t) F(v).$$

Therefore,

$$\int_{V_A(t)} \max_i [p_i F(v|i)] dv \geq (1-t) \int_{V_A(t)} F(v) dv$$

which, with (7') and (8'), is

$$C(t) \geq (1-t)\, A(t).$$

Hence

$$E(t) \leq tA(t) \leq t .$$

(c) is shown in the following section.

Error-Reject Tradeoff

A complete description of the performance of recognition systems is given by the error-reject tradeoff, i.e., the functional relation of E and R at all levels. A typical tradeoff curve is given in Fig. 2. Since both E and R of the optimum recognition systems are monotonic functions of the rejection threshold t, one can compute the tradeoff E vs. R from E(t) and R(t).

We shall now show that the rejection function R(t) alone suffices to completely characterize the optimum recognition performance. In other words, E can be derived from R(t), or from its inverse t(R). The central result is the simple functional relation between E and R, namely

$$E = \int_R^1 t(R)dR. \tag{13}$$

This relation is valid for all optimum decision rules as defined in Equations (1) - (4). No explicit forms for the density functions $F(v|i)$ are required in deriving the integral relation of (13). However, it will be assumed for convenience in the following derivation that R(t) is differentiable with respect to t. Under this assumption, the inverse function t(R) is single-valued. However, this assumption will later be removed.

Consider an decremental change in the rejection threshold from t to t - $\Delta t$; the reject region expands from $V_R(t)$ to $V_R(t-\Delta t)$. Let $\Delta V_R(t)$ denote the incremental region $V_R(t-\Delta t)$ - $V_R(t)$. For any v in $\Delta V_R(t)$, it was accepted at the threshold t and is now rejected at the lower threshold t - $\Delta t$. Equations (2) and (4) now give:

$$(1-t)F(v) \leq \max_i p_i F(v|i) < (1-t+\Delta t)F(v) \text{ for } v \in \Delta V_R(t) \tag{14}$$

By integrating the last expression over the incremental region $\Delta V_R$, one obtains

$$(1-t)\,\Delta R \leq -\Delta C < (1-t+\Delta t)\,\Delta R \tag{15}$$

where $\Delta R$ and $\Delta C$ are respectively the increments in the rejection rate and correct recognition rate, namely

$$\Delta R = \int_{\Delta V_R} F(v)dv$$

$$\Delta C = \int_{\Delta V_R} \max_i [p_i F(v|i)]dv.$$

Of course, the increment in the error rate is simply

$$\Delta E = -\Delta R - \Delta C.$$

By substituting (16) into (15), one has

$$-t\Delta R \leq \Delta E < -(t-\Delta t)\,\Delta R \tag{17a}$$

$$-t\,\frac{\Delta R}{\Delta t} \leq \frac{\Delta E}{\Delta t} < -t\,\frac{\Delta R}{\Delta t} + \Delta R. \tag{17b}$$

Since R(t) is differentiable, $\Delta R \to 0$ as $\Delta t \to$ ) and (17) yields

$$\frac{dE}{dt} = -t\,\frac{dR}{dt}\,. \tag{18}$$

By integrating (18) from t = 0 to t, one has

$$E(t) - E(0) = -\int_0^t t\,\frac{dR}{dt}\,dt$$

$$= -\int_{R(0)}^{R(t)} t(R)\,dR\,.$$

Since E(0) = 0 and R(0) = 1, the above expression becomes

$$E = \int_R^1 t\,dR\,. \tag{13}$$

This relation is depicted in Fig. 3. Equation (13) can also be written, through an integration by parts and as indicated in Fig. 3, as

$$E = \int_o^t R(t)\,dt - tR(t). \tag{19}$$

Equation (18) gives

$$\frac{dE}{dR} = -t \leq 0 \tag{20}$$

The rejection threshold is the differential error-reject tradeoff. In particular, the initial slope of the error-reject curve is $-1 + \frac{1}{n}$ or greater while the final slope is 0. Equation 20 also gives:

$$\frac{d^2E}{dR^2} = -\frac{dt}{dR} \geq 0 \tag{21}$$

The optimum error-reject curve is always concave upward and the slope increases from -1 to 0 as R increases from 0 to 1. (Figure 2)

Although the integral relation (13) is derived under the assumption that R(t) is differentiable, the assumption of differentiability is not essentiable. For example, it suffices to assume that R(t) is continuous. A proof for (13) would start with (17a) instead of (17b). Actually no assumption about R(t) is necessary for the validity of (13). R(t) is, of course, monotonic and bounded function of t and is thus of bounded variation. Consequently, the Stieltjes integral $\int_{t=0}^{t} t \; dR(t)$ always exists. The error-reject integral is in general

$$E(t) = -\int_{t=0}^{t} t\, dR(t) \qquad (22)$$

To show (22) we first sum (17a) with t steadily increasing throughout the range of interest to obtain

$$-\sum t\,\Delta R \leq E(t) \leq -\sum t\Delta R - \sum \Delta t \Delta R$$

and then let $\Delta t$ tend to zero. As $\Delta t$ tends to zero the last sum of the above expression vanishes and (22) results.

Another Proof of the Error-Reject Integral

We shall now present another proof* of the error-reject integral of (13) with a hope to provide additional insight to the relation. Let M(v) denote the random variable:

$$M(v) = \frac{\max_i p_i F(v|i)}{F(v)} \tag{23}$$

M(v) is the maximum of the a posteriori probabilities of the classes given the pattern v. Let g(m) be the density function of M. In general g(m) is rather complex function of the underlying density functions $F(v|i)$, however, its explicit form does not concern our proof here. In terms of the variable M(v), the reject condition, Eq. (4) becomes

$$M(v) \leq 1 - t \tag{24}$$

and the probability of reject is:

$$R(t) = \int_0^{1-t} g(m)\, dm \tag{25}$$

which also gives

$$g(m) = \frac{dR(1-m)}{dm} \tag{26}$$

*This proof is due to Mr. M. Hellman and Dr. J. Raviv of IBM Watson Research Center.

Since g(m) is non-negative, R(t) is a monotonic increasing function of the upper limit of the integral of (25), or a monotonic decreasing function of t. By the definition of M(v), the probability of correct recognition for a given reject threshold t is

$$C(t) = \int_{1-t}^{1} m g(m) dm \tag{27}$$

By substituting (26) in (27) and integrating by parts, we obtain:

$$C(t) = \int_{t}^{0} (1-t)\, dR(t)$$

$$= 1 - (1-t)R(t) - \int_{0}^{t} R(t)dt. \tag{28}$$

The error rate is then

$$E(t) = 1 - R(t) - C(t)$$

$$= \int_{0}^{t} R(t)dt - tR(t).$$

which is (19) and is equivalent to (13).

Rejection Threshold of a Minimum Risk Rule

It is known [1] that the optimum decision rule given in Equations (1) to (4) is also a minimum risk rule if the cost function is uniform within each class of decisions, i.e. if no distinction is made among the errors, among the rejects, and among the correct recognition. The rejection threshold is then related to the costs as follows:

$$t = \frac{W_r - W_c}{W_e - W_c} \tag{29}$$

where $W_e$, $W_r$, and $W_c$ are the costs for making an error, reject and correct recognition respectively. Usually $W_e > W_r > W_c$. The rejection threshold is simply the normalized cost for the rejection. We can take $W_c = 0$ and $W_e = 1$, and the minimum risk is

$$\text{Risk}(t) = E(t) + tR(t) = \int_0^t R(t)\, dt \tag{30}$$

which is also depicted in Fig. 3.

Examples

For numerical illustration, two examples are given here. In these examples, the pattern vector v is one-dimensional and there are two pattern classes with equal apriori probability ofoccurrence, i.e. $p_1 = p_2 = \frac{1}{2}$. The examples respectively are concerned with the normal distributions and uniform distributions.

For two classes, the condition for rejection, namely Eq. (4), can never be satisfied when $t > \frac{1}{2}$, hence the reject rate is always zero if the reject threshold, t, exceeds $\frac{1}{2}$. The effective range of t for two classes of problems is, therefore, from 0 to $\frac{1}{2}$. With n = 2 and $0 \leq t \leq \frac{1}{2}$, it can be readily verified that the condition for rejection, (4) is equivalent to

$$\frac{t}{1-t} \leq \frac{p_1 F(v|1)}{p_2 F(v|2)} \leq \frac{1-t}{t} \tag{31}$$

(1) Normal Distributions of Equal Variance

Consider two normal distributions with means $\mu_1$ and $\mu_2$ and equal covariance $\sigma^2$ (Fig. 4). Take $\mu_1 > \mu_2$. The density function is, i = 1 or 2,

$$F(v|i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(v-\mu_i)^2}{2\sigma^2}\right] \tag{32}$$

With (31) and some algebraic manipulations, (32) can be transformed to

$$\left| v - \frac{1}{2}(\mu_1 + \mu_2) \right| \leq \frac{\sigma^2}{\mu_1 - \mu_2} \ln\left(\frac{1}{t} - 1\right) \qquad (33)$$

i.e., the optimum rule is to reject whenever the pattern lies within a certain distance of the midpoint between the two means. The corresponding error and reject rates are:

$$E(t) = \Phi(a) \qquad (34a)$$

$$R(t) = \Phi(b) - \Phi(a) \qquad (34b)$$

where $\Phi$ is the normal cumulative distribution function, namely

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}} dx \qquad (35)$$

and the parameters are

$$a = -\frac{1}{2}s - \frac{1}{s}\ln\left(\frac{1}{t} - 1\right) \qquad (a)$$

$$b = -\frac{1}{2}s + \frac{1}{s}\ln\left(\frac{1}{t} - 1\right) \qquad (b) \quad (36)$$

$$s = (\mu_1 - \mu_2)/\sigma \qquad (c)$$

The parameter $s$ is the (normalized) separation between the means of the distributions and is the only (composite) parameter of the distributions that $R(t)$ and $E(t)$ depend upon. It is straightforward to verify (18) and hence (13) for this example. A set of the error, reject, and tradeoff curves (for $s = 1, 2, 3,$ and $4$) is depicted in Fig. 5.

2. Uniform distributions

Consider two uniform distributions:

$$F(v|1) = \begin{cases} 1 & \text{when } 0 \le v \le 1 \\ 0 & \text{elsewhere} \end{cases} \tag{37a}$$

$$F(v|2) = \begin{cases} \frac{1}{2} & \text{when } \frac{1}{2} \le v \le \frac{5}{2} \\ 0 & \text{elsewhere} \end{cases} \tag{37b}$$

which are shown in Fig. 6.

The reject function $R(t)$ is simply:

$$R(t) = \begin{cases} \frac{3}{8} & \text{when } 0 \le t \le \frac{1}{3} \\ 0 & \text{when } \frac{1}{3} < t \le 1 \end{cases} \tag{38}$$

which is discontinuous (Fig. 7a) and the integral of (21) is evaluated to

$$E(t) = \begin{cases} 0 & \text{when } 0 \le t \le \frac{1}{3} \\ \\ \frac{1}{8} & \text{when } \frac{1}{3} < t \le 1 \end{cases} \tag{39}$$

which is shown in Fig. (7b) and the tradeoff can assume only two values, namely $(E, R) = (\frac{1}{8}, 0)$ or $(0, \frac{3}{8})$ (Fig. 7c). However, if a randomized scheme is used in the range $\frac{1}{2} \le v \le 1$, R may vary continuously from $\frac{1}{8}$ to 0 as shown by the dotted line in Fig. (7c).

Some Practical Implications

Most of our results on the error-reject tradeoff seem consistent with our intuition, although the simple integral relation between the error and reject rates is somewhat unexpected. These results have some practical implications are are useful in system design and performance evaluation.

Since the slope of the error-reject tradeoff curve (Fig. 2) is the value of the rejection threshold, the tradeoff is most effective initially (i.e. at the low level of rejection) and it gets harder as the error rate is lower. This is certainly common in our practical experience; excessive rejection is generally required to reduce residual errors.

Practical applications of the present results are in the areas of system design and performance evaluation of the recognition systems. The general characteristics of the error-reject tradeoff curve provide the system designer a convenient means of verifying the basic assumption on the underlying probability distributions. The integral of (13) makes it possible to calculate error rates, and consequently the tradeoff curve from the empirically observed reject rates. No class identification of the sample patterns are required in obtaining the empirical rejection curve. Or equivalently one can just obtain an empirical density

function is the maximum of the a posteriori probabilities, and then calculate the error and reject rates using equations (25) and (27).

In most recognition tasks, the underlying probability distributions of the patterns are not completely known and the design of the recognition systems is generally based on empirical data. A common design procedure is to assume, on the basis of available (usually limited) a priori information and the designer's intuition, some functional forms of the distributions and to derive the system structure based on these assumptions and to adjust the system parameters by using the empirical data. It is not always a simple matter to verify the validity of the assumptions on which the system structure is based. However, one can always, though laboriously, obtain the empirical error-reject tradeoff curve and compute the theoretical one from the basic assumptions. A comparison of the empirical and theoretical tradeoff curves can quickly reveal how well the theoretical model agrees with the empirical data and serve as a checkpoint for initiating the process of revising and improving the theoretical model.

The data used in any meaningful evaluation of a recognition system is usually large and it is extremely costly and

## Conclusion

A general error and reject tradeoff relation is derived for the (Bayes) optimum recognition system with an option to reject. The error probability is a Stieltjes integral of the rejection threshold with respect to the reject probability. The error function can be directly evaluated from the reject function. Hence, the reject function determines the recognition error and reject tradeoff and completely characterizes the performance of the optimum recognition system.

Some practical implications in the system design and performance evaluation of the recognition systems are discussed. The error-reject integral provides a simple means of calculating the error rate from the empirical reject curve without actually identifying the recognition errors.

time consuming to detect the recognition errors. To identify a recognition error, additional information usually human inspection at some stage, is required. On the other hand, the rejection is the explicit result of a definite decision, and the rejects can be readily recorded and tallied. Equation (13) provides a simple means of calculating the error rate from the reject curve without actually identifying the errors.

## REFERENCES

1. C. K. Chow, "An Optimum Character Recognition System Using Decision Functions", IRE Trans. on Electronic Computers, Vol. EC-6, pp. 247 - 254, December, 1957.

2. C. K. Chow and C. N. Liu, "An Approach to Structure Adaptation in Pattern Recognition", IEEE Trans. Systems Science and Cybernetics, Vol. SSC-2, pp. 73-80, December, 1966.

3. R. Bakis, N. Herbst and G. Nagy, "An Experimental Study of Machine Recognition of Hand-printed Numerals", IEEE Transactions Systems, Science and Cybernetics, Vol. SSC-4, pp. 119-132, July, 1968.

4. R. B. Hennis, "The IBM 1975 Optical Page Reader, Part I: System Design", IBM Journal of Research and Development, Vol. 12, pp. 346 - 353, September, 1968.

This page is missing from the original document.

## FIGURE CAPTIONS

1. Reject Regions in the Pattern Space

2. Error-Reject Tradeoff Curve

3. Reject Curve

4. Example in Normal Distribution

5. Normal Distributions: (a) Reject and Error Curves (b) Tradeoff Curve

6. Example in Uniform Distributions

7. Uniform Distributions: (a) Reject Curve (b) Error Curve (c) Tradeoff Curve

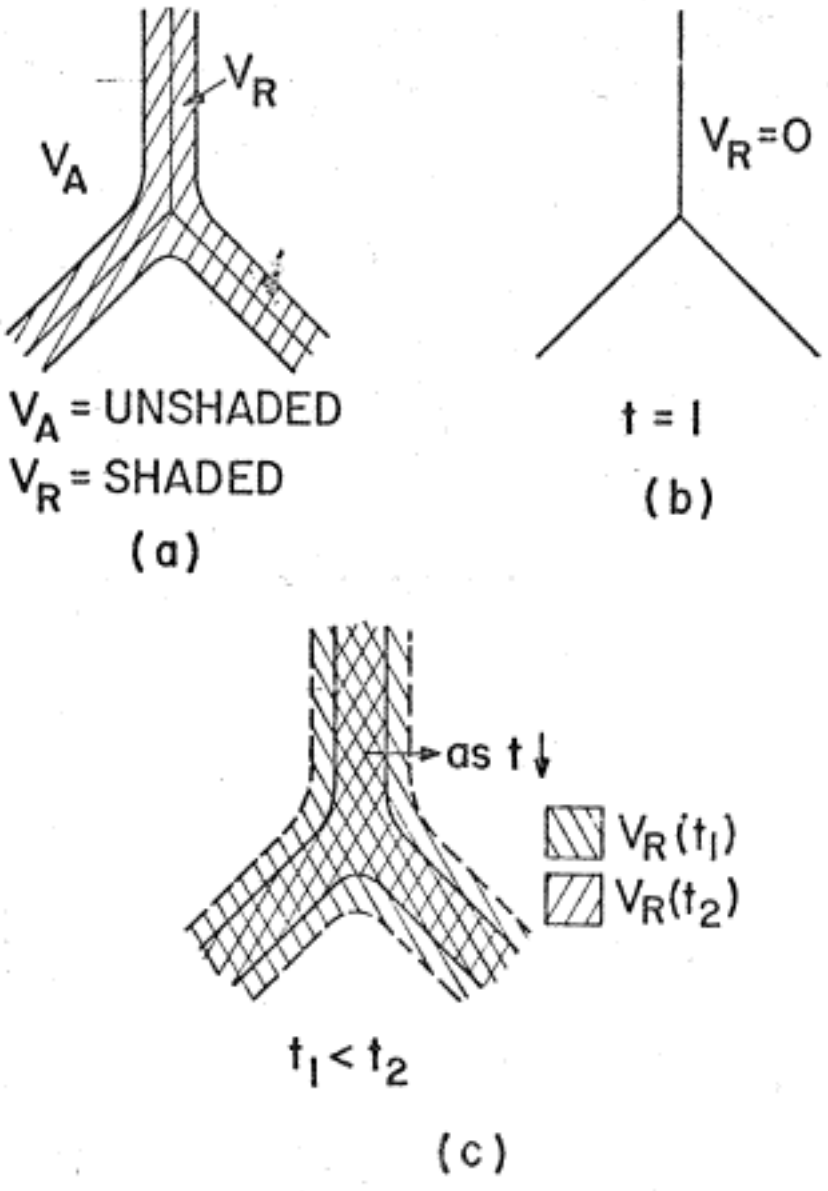


Figure 1

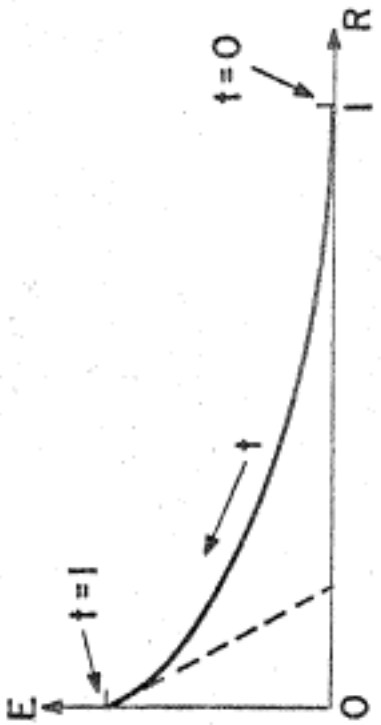


Figure 2

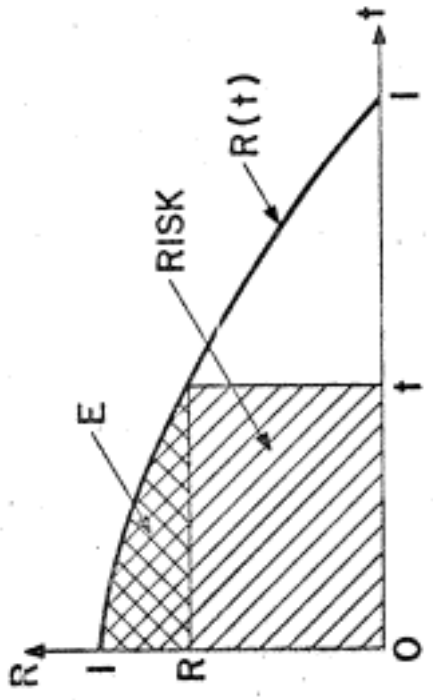


Figure 3

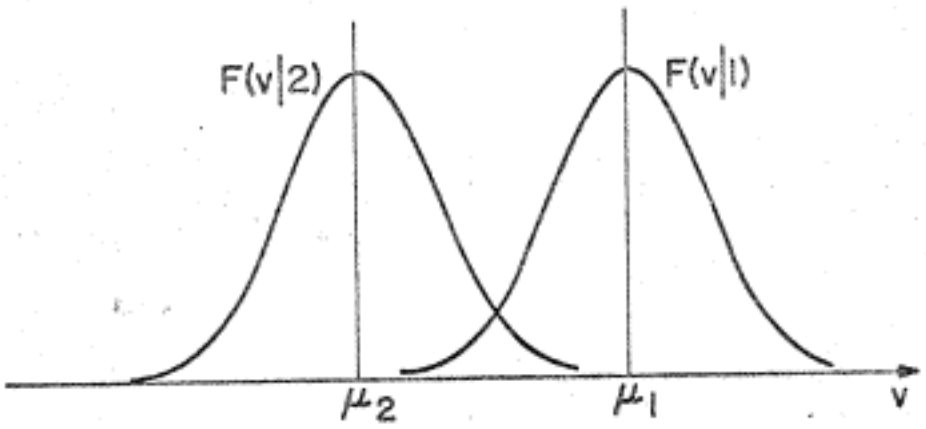


Figure 4

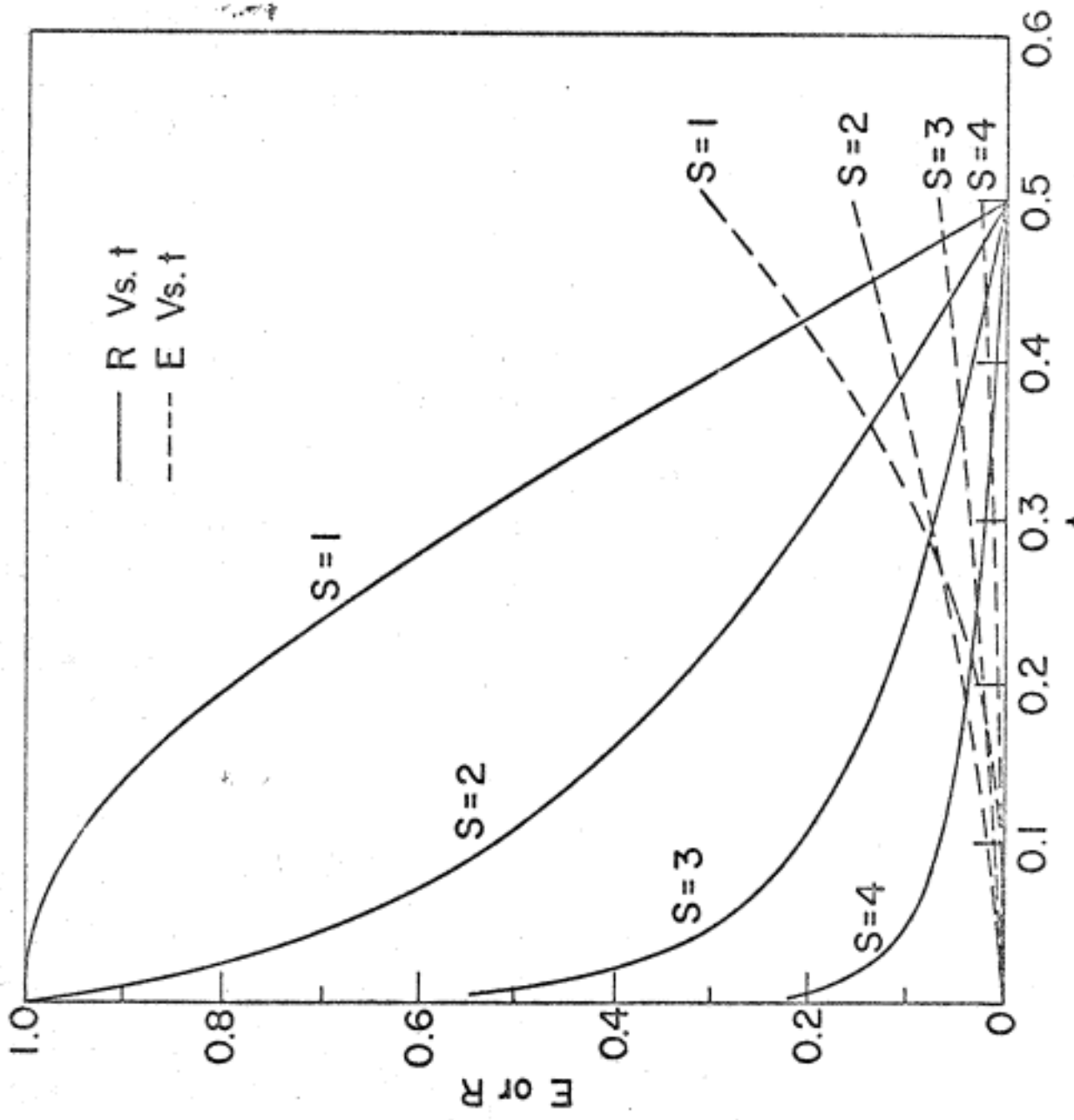
R Vs. t
E Vs. t
S=1
S=2
S=3
S=4
S=1
S=2
S=3
S=4
E or R
t
0
0.2
0.4
0.6
0.8
1.0
0.1
0.2
0.3
0.4
0.5
0.6

Figure 5(a)

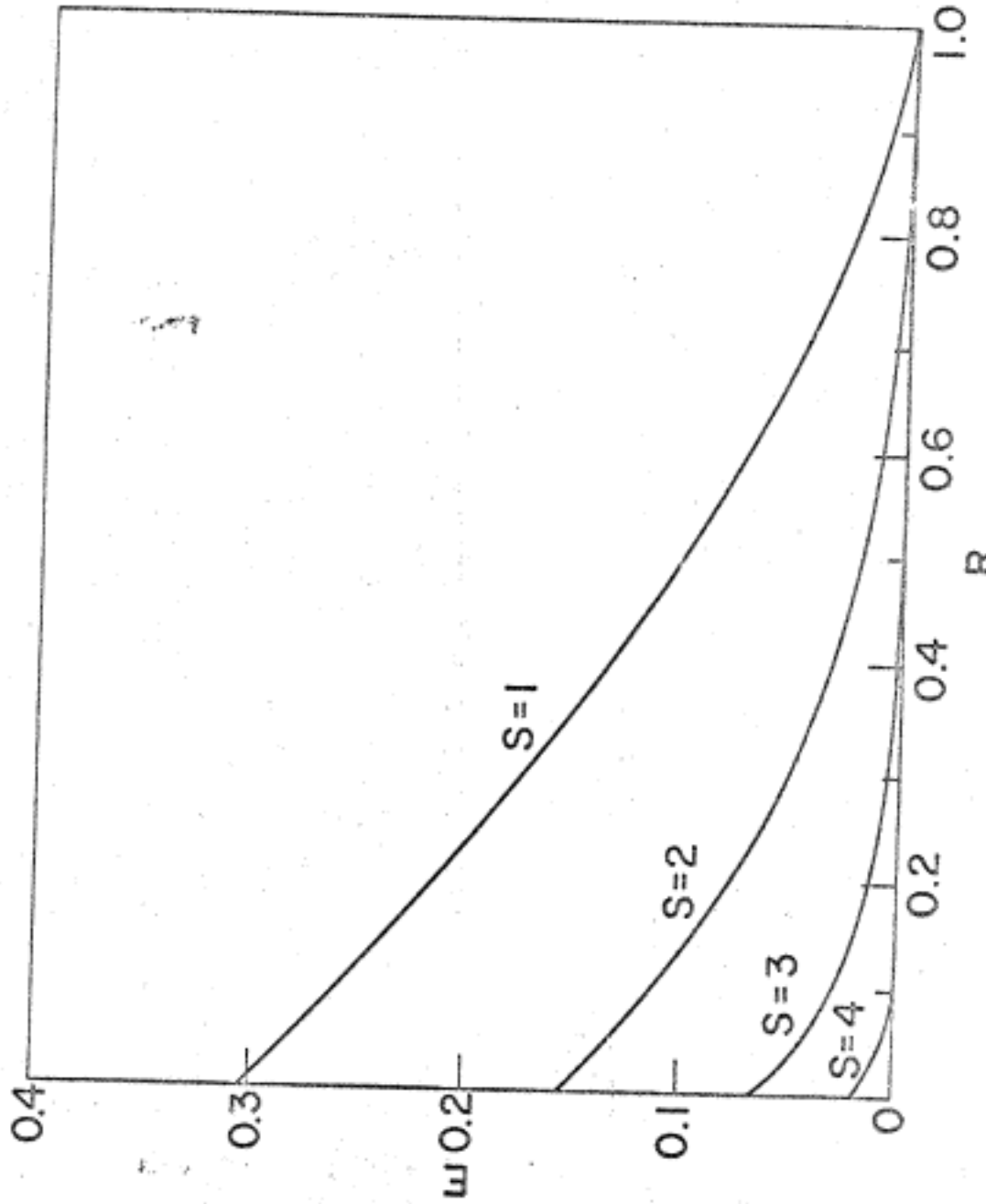


Figure 5(b)

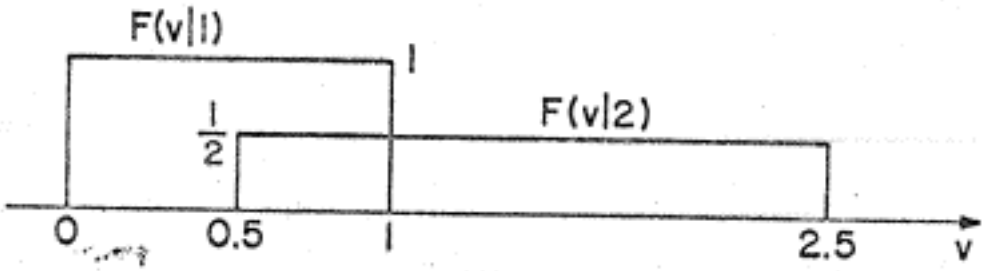


Figure 6

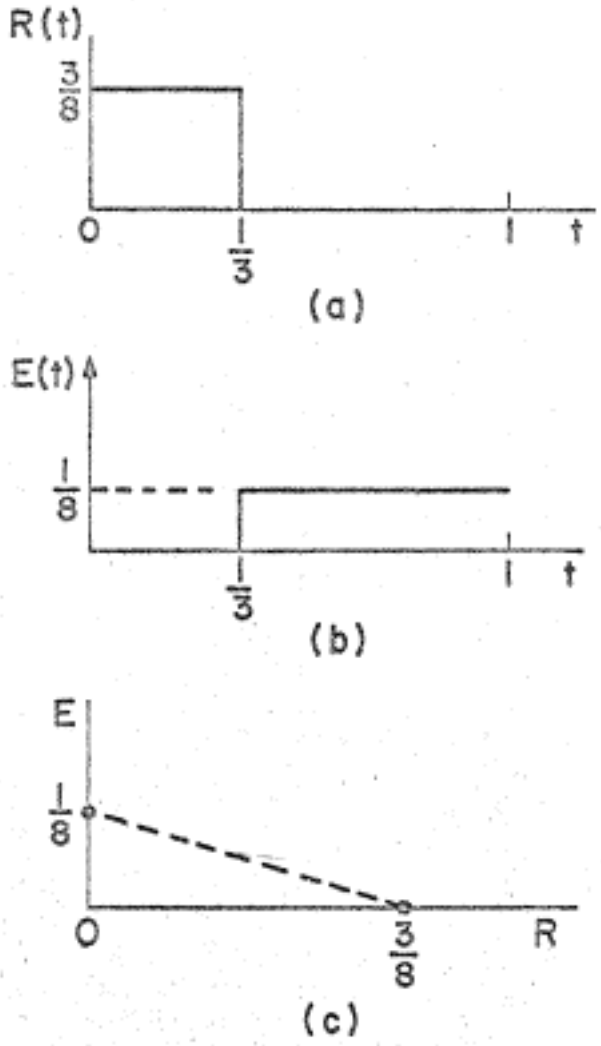


Figure 7