MAA PROGRAM FOR MATHEMATICS AND THE BEHAVIORAL SCIENCES

Psychology

Lecturer: R. Duncan Luce

Notes by: Henry A. Selby

In this section of the seminar we shall orient the re-
marks around aspects of two substantive topics: preference
and psychophysics. We shall present both algebraic and
probabilistic formulations with special emphasis on conditions
under which numbers can be introduced into the analysis of
social scientific data, as well as in theory construction. We
begin with one problem in the foundations of measurement in
the social (and physical) sciences.

Foundations of Measurement*

One fundamental mathematical result which is a basis for
Hölder's theorem states that any Archimedean ordered group is
homomorphic to a subgroup of the additive real numbers, i.e.,
there exists a function from the given group into the reals
such that

---

*The reference for this first section of the lectures
is Krantz, David, R. Duncan Luce, Patrick Suppes and Amos
Tversky (1972) Foundations of Measurement. New York Academic
Press. In particular the reader is referred to chapters 1,2,
3,6,7, and 9. See also Krantz, D. H. (1968) "A Survey of
Measurement Theory" in Dantzig, G. B. and A. T. Veinott (eds).
Mathematics of the Decision Sciences, II. Providence, R. I.
American Mathematical Society, pp. 314-350; Pfanzagl, J.
(1968) Theory of Measurement. New York, Wiley; and Fishburn,
P. C. (1970) Utility Theory and Decision Making, New York, Wiley.

MAA PROGRAM FOR MATHEMATICS AND THE BEHAVIORAL SCIENCES

Psychology

Lecturer: R. Duncan Luce

Notes by: Henry A. Selby

In this section of the seminar we shall orient the re-
marks around aspects of two substantive topics: preference
and psychophysics. We shall present both algebraic and
probabilistic formulations with special emphasis on conditions
under which numbers can be introduced into the analysis of
social scientific data, as well as in theory construction. We
begin with one problem in the foundations of measurement in
the social (and physical) sciences.

Foundations of Measurement*

One fundamental mathematical result which is a basis for
Hölder's theorem states that any Archimedean ordered group is
homomorphic to a subgroup of the additive real numbers, i.e.,
there exists a function from the given group into the reals
such that

---

*The reference for this first section of the lectures
is Krantz, David, R. Duncan Luce, Patrick Suppes and Amos
Tversky (1972) Foundations of Measurement. New York Academic
Press. In particular the reader is referred to chapters 1,2,
3,6,7, and 9. See also Krantz, D. H. (1968) "A Survey of
Measurement Theory" in Dantzig, G. B. and A. T. Veinott (eds).
Mathematics of the Decision Sciences, II. Providence, R. I.
American Mathematical Society, pp. 314-350; Pfanzagl, J.
(1968) Theory of Measurement. New York, Wiley; and Fishburn,
P. C. (1970) Utility Theory and Decision Making, New York, Wiley.

$$\phi(aob) = \phi(a) + \phi(b)$$

$$a \leq b \quad \text{iff} \quad \phi(a) \leq \phi(b).$$

The axioms of an Archimedean ordered group will be given implicitly below. This simple axiom is an idealized account of the measurement of mass length etc., in the physical sciences, but it has not been useful in the social sciences. It seems not possible to find suitable concatenation (group or semi-group) operations in the social sciences. Nonetheless, intuitively it appears that there ought to be some measureable scales: utility in economics, loudness and pitch of tones and brightness of lights is psychophysics, and many others such as hunger, intelligence, etc., in psychology. All of these seem to have numerical properties. Consider the loudness example from psychophysics. Clearly loudness of a pure tone is not the same as intensity (amplitude of the sinusoid) -- the latter spans thirteen orders of magnitude, but not the former. Moreover what is judged as a reduction by half in loudness is not half in intensity.

One recently proposed system of measurement does not require any explicit concatenation operation. It certainly applies to some physical measurement and it appears potentially useful in the social sciences. Instead of a concatenation operation on a set, the basic set is assumed to be a Cartesian product of factors and there is an ordering over this product set. In particular, suppose there are n sets A_i, i = 1,...,n, then the ordering is on $\underset{i=1}{\overset{n}{X}} A_i$. A physical

example is the ordering by momentum when the factors are the mass and velocity of objects. In one psychophysical example let A_1 be the set of intensities of a pure tone presented to one ear, and A_2 those to the other ear, and let the ordering of $A_1 \times A_2$ be judged overall loudness (see Krantz et al. 1972: Chapter 2 and p. 269).

The problem of finding a suitable axiomatic set can be stated as follows. We require axioms that will lead to a function F, where, F: Re X Re → Re, and the two functions ϕ_i, where ϕ_i: A_i → Re, such that for all a,b ε A_1, and p,q ε A_2

ap $\gtrsim$ bq   iff $F[\phi_1(a), \phi_2(p)] \geq F[\phi_1(b), \phi_2(q)]$.

The special case of this general representation that we shall study is the additive one, i.e.,

$$F(x,y) = x + y.$$

We desire axioms sufficient to lead to such an additive representation. One set is:

(1) The ordering must be a <u>weak</u> ordering, i.e., given any pair a,b, then either a $\gtrsim$ b, or b $\gtrsim$ a (connectedness) and $\gtrsim$ is transitive.

(2) Independence. For all a,b ε A_1, and p,q ε A_2

ap $\gtrsim$ bp   iff   aq $\gtrsim$ bq

and   ap $\gtrsim$ aq   iff   bp $\gtrsim$ bq .

From this we may induce orderings on A_i as follows:

a $\gtrsim_1$ b if there exists p ε A_2 such that ap $\gtrsim$ bp, p $\gtrsim_2$ q if there exists a ε A_1 such that ap $\gtrsim$ aq. From the axiom, the

ordering on A_1 is independent of the choice of p and the ordering on A_2 is independent of the choice of a.

(3) Double Cancellation. For all a,b,f ϵ A_1, p,q,x ϵ A_2, if ax $\gtrsim$ fq and fp $\gtrsim$ bx, then ap $\gtrsim$ bq. (This property is called the Thomsen condition in the theory of webs.)

(4) The Archimedean Property. The ordering must exhibit the qualitative property which is the analogue of the usual Archimedean property of numbers, namely that all non-zero numbers are commensurate. (For an explicit statement of this axiom see Krantz et al 1972:263.)

(5) Unrestricted Solvability. Given any three of a,b, ϵ A_1, p,q, ϵ A_2, the fourth exists, such that ap $\gtrsim$ bq and bq $\gtrsim$ ap, which we abbreviate as ap $\sim$ bq.

Theorem

If axioms 1 - 5 hold, then there exist functions ϕ_i: $A_i \to$ Re such that

$$ap \gtrsim bq \quad \text{iff} \quad \phi_1(a) + \phi_2(b) \gtrsim \phi_1(b) + \phi_2(q)$$

If ϕ_i exhibit the same property then there exists $\alpha > 0$, and β_i that

$$\phi_i = \alpha\phi_i + \beta_i \ .$$

Outline of Proof of Theorem

(Throughout the proof we use unrestricted solvability without explicit mention.)

Choose any a_0 ϵ A_1, p_0 ϵ A_2 and define α_0, π, o_1 as solutions of the following indifferences:

$$a_0\pi(a) \sim ap_0, \quad \alpha(p)p_0 \sim a_0p_1, \quad (ao_1b)p_0 \sim a\pi(b).$$

Intuitively, if we represent the factors as lines, then with a_o, p_o and a given $\pi(a)$ is selected so that combined with a_o it just balances the effect of a combined with p_o as shown in the sketch.
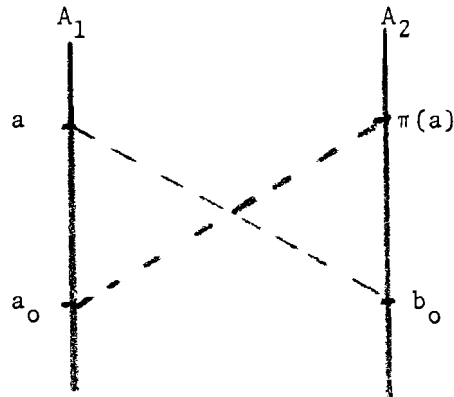
The following statements are easily proved:

$$a\pi(b) \sim b\pi(a)$$

$$ao_1 b \underset{-1}{\sim} bo_1 a$$

$$a_o \pi(a_o) \sim a_o p_o \quad \text{(which implied } \pi(a_o) \sim_2 p_o)$$

$$(p)[\alpha(p)] \sim_2 p$$

$$(a)[\pi(a)] \sim_1 a$$

Using these we show that the system A_1, o_1, $\sim_1$ is in fact an Archimedean ordered group.

1. a_o is an identity

By definition $(ao_1 a_o)p_o \sim a\pi(a_o) \sim ap_o$, whence by definition of $\gtrsim_1$ $ao_1 a_o \sim_1 a$.

2. Existence of Inverses. Let p solve $ap \sim a_o p_o$ and define $a^{-1} = \alpha(p)$. Then $(aoa^{-1})p_o \sim a\pi(a^{-1}) \sim a\pi[\alpha(p)] \sim ap \sim a_o p_o$,

and so $ao_1a^{-1} \sim_1 a_o$. The proof that $a^{-1}o_1a \sim_1 a_o$ is similar.

3. Associativity. Observe $b\pi(\pi) \sim (ao_1b)p_o$ and $(bo_1c)p_o \sim b\pi(c)$
imply, by Axiom 3, $b(o_1c)\pi(a) \sim (ao_1b)\pi(c)$. Thus,

$$[ao_1(bo_1c)]p_o \sim a\pi(bo_1c)$$
$$\sim (bo_1c)\pi(a)$$
$$\sim (ao_1b)\pi(b)$$
$$\sim [ao_1(bo_1c)]p_o$$

4. Monotonicity.

$$a \geq_1 b \quad \text{iff} \quad a\pi(c) \succsim_1 b\pi(c)$$
$$\text{iff} \quad (ao_1c)p_o \succsim_1 (bo_1c)p_o$$
$$\text{iff} \quad ao_1c \succsim_1 bo_1c$$

5. The Archimedean property follows from the Archimedean
axiom, (no detailed proof is given). By Hölder's theorem
there exists an additive homomorphism ϕ_1 of A_1. Define
$\phi_2(p) = \phi_1\alpha(p)$. Next we note that $ap \succsim bq$ iff

$ao_1 (p) \geq_1 bo_1\alpha(q)$. Since from $a_op \sim (p)p_o \sim a_o\pi[\alpha(p)]$
independence implies $ap \sim a\pi[\alpha(p)] \sim [a_oo_1\alpha(p)]$ p_o.
Similarly, $bq \sim [b_oo_1\alpha(q)]$ p_o. The result follows by in-
dependence. Thus, $ap \succsim bq$ iff $\phi_1(a) + \phi_1\alpha(p) \geq \phi_1(b) +$
$o_1[\alpha(q)]$ iff $\phi_1(a) + \phi_2(p) \geq \phi_1(b) + \phi_2(q)$.

The uniqueness follows from that of Hölder's result: we
do not prove it explicitly.

Note that the solvability property is a very strong
constraint which is not often satisfied in social science
situations, nor, for that matter in practical physics. For
example, a plot of equal loudness contours as a function of

intensity and frequency of pure tones is shown in Figure 1,

Figure 1

Equal Loudness Contours as Function of Intensity & Frequency

Frequency
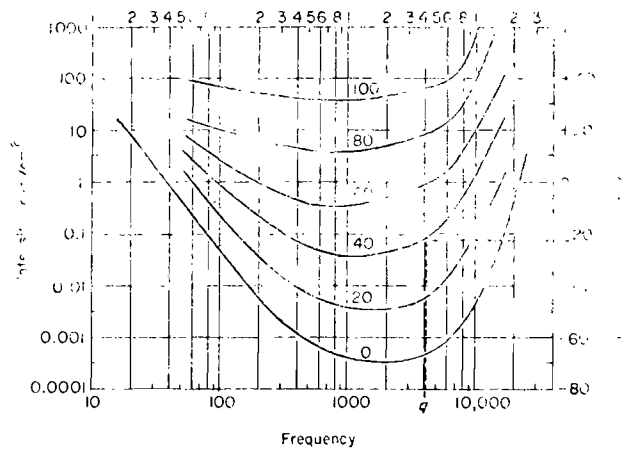
and we see that ap ∿ bq is not always solvable. (See Krantz et al. 1972:255-6 for full example.) The continuity of the components makes clear that ap ∿ bq has a solution if and only if there exist $\bar{b}$, $\underline{b}$ ε A_1 such that $\bar{b}q \gtrsim ap \gtrsim \underline{b}q$; a similar statement holds for the second component. An axiomatization involving this form of restricted solvability is given in Chapter Six of Krantz et al.

A second assumption is also troublesome: the Archimedean axiom. Louis Nerans (University of California, Irvine) has worked various axiomatizations without postulating the Archimedean axiom; he develops a representation into non-

standard arithmetic, and not the real numbers

Generalization to n Factors

For $n \geq 3$ factors, the axiomatization is rather simpler than in the two-factor case. A general independence notion replaces both two-factor independence and double cancellation. If one takes an arbitrary subset of the factors it is assumed that the ordering induced on the product of these factors, with the elements of the remaining factors held fixed, is independent of that choice of fixed elements. This property implies double cancellation on each pair of components. The $n = 2$ result permits one to construct numerical representations on each pair of factors, which then in turn can be shown to serve to give a general additive representation in the n-dimensional case. The details can be found in section 6.11 and 6.12 of Krantz et al.

Application to Expected Utility Theory

Conjoint measurement can be applied to the problem of calculating expected utility, which has to do with the ordering among gambles with uncertain outcomes. A gamble, or better, conditional decision can be defined as follows: Let A be an event (such as the throw of a die), then the decision is a function mapping A into some set C of consequences, i.e., $f_A: A \to C$.

Suppose the set of all conditional decisions is ordered by preference relation $\succsim$. The following question is posed in the theory of expected utility: under what conditions on

$\langle \mathcal{D}, \gtrsim \rangle$ does a probability function P on events and a utility function u on $\mathcal{D}$ exist such that

(i) $f_A \gtrsim g_B$ iff $u(f_A) \geq u(g_B)$,

(ii)   if $A \cap B = \phi$, $u(f_A \cup g_B) = u(f_A)p(A|A \cup B) +$

$$u(g_B)p(B|A \cup B).$$

Property (ii) is known as the conditional expected utility property. In Chapter eight of Krantz et al. a reasonably compact set of axioms is given which permit this construction, with $\mathcal{D}$ restricted to a subset of all such functions, but subject to closure assumptions. The proof of the result rests on the theorem for conjoint measurement. A rough indication of its role is given. For each integer n, let $\{A_i\}$, $i \in 1,\ldots,n$ be any set of pairwise disjoint events. Let $\mathcal{D}_{A_i}$ be the set of all functions f_{A_i} . And let the order $\gtrsim'$ on $\overset{n}{\underset{i=1}{X}}$ be defined as follows: for f_{A_i}, $g_{A_i} \in \mathcal{D}A_i$,

$$(f_{A_1}, f_{A_2}, \ldots, f_{A_n}) \gtrsim' (g_{A_1}, g_{A_2}, \ldots, g_{A_m})$$

iff $f_{A_1} \cup f_{A_2} \cdots \cup f_{A_n} \gtrsim g_{A_1} \cup g_{A_2} \cup \cdots \cup g_{A_m}$

One uses assumed properties of $\gtrsim$ to show that

$\left\langle \overset{n}{\underset{i=1}{X}} \mathcal{D}_{A_i}, \gtrsim' \right\rangle$ satisfies the axioms of n-dimensional additive

conjoint measurement, and so there is an additive representation. Each set A appears in many such substructures. Let Φ_A denote the set of all numerical functions that arise from all such structures. A non-trivial step in the proof involves

showing that any two $\phi_1\phi'$ ϵ Φ_A are related by a positive linear transformation.

Select any f_o, f_1 ϵ $\mathcal{D}$ such that $f_1 > f_o$. For every $A \neq \phi$, use solvability to select decisions $u(a)[n_j(A)] = j$, which is possible by the free linear transformation available. For any two disjoint events A and B, let $\phi_{A,B}$ + $\phi_{B,A}$ denote the additive representation of $\langle \mathcal{D}_A \times \mathcal{D}_B, \gtrsim' \rangle$ normalized so that $\phi_{A,B}$ + $\phi_{B,A}$ = $u_{A \cup B}$. But by the property of Φ_A, Φ_B, there exist constants, which we write as $P(A|A \cup B)$, $P(B|A \cup B)$, and $\beta_{A_1 B}$ and $\beta_{B_1 A}$ so that

$$\phi_{A,B} = P(A|A \cup B)u_A + \beta_{A_1 B}$$

$$\phi_{B,A} = P(B|A \cup B)u_B + \beta_{B_1 A} .$$

The final step in the proof involves using the assumptions to show $\beta_{A_1 B}$ + $\beta_{B_1 A}$ = 0 and $P(A|A \cup B)$ + $P(B|A \cup B)$ = 1.

An example of an empirical test of the properties of the conjoint measurement system can be taken from the work of Levelt and his colleagues (1971)[*]. The account is taken from Krantz et al. (1972:269)

"In a study of binaural loudness, Levelt, Riemersma and Bunt (1971) formed 36 stimuli by presenting all possible

---

[*]Levelt, W.J.M., Reimersma, J.B., and Bunt, A.A. "Binaural Additivity of Loudness." Technical Report NR:HB-71-70EX, R.U. Groningen, Netherlands, Heymans Pulletins Pschologische Instituten. 1971

combinations of one of six different 1000-Hz signals to each
of the ears. The signals were in equal-decibel steps from
20 to 70 dB Sound Pressure Level (SPL). Two subjects ordered,
according to loudness those pairs for which the ordering is
not predictable by independence. Using a computer program,
a best fitting additive representation was constructed. To
an excellent first approximation, it was found to be of the
form

$$\alpha_r I_r^{\beta_r} + \alpha_l I_l^{\beta_l} + \gamma$$

where I_r, I_l are the physical intensities to the right and
left ears. Not only does this say that the data are approxi-
mately additive, but that each additive scale is a power
function of intensity. Moreover the dependence is different
for the two ears. Combining the above data with other re-
lated data, the pairs of exponents were 0.44, 0.61 and 0.41,
0.47 for these subjects."

The experiment was motivated by the theory of conjoint
measurement in an attempt to see whether the structure of
the system could be satisfied.

Models of Choice

The problem of error in measurement is chronic in any
science, but is particularly acute in the social ones. It
is very difficult for social scientists to fit algebraic or
analytic models to their data. In addition to attempting
to increase sample sizes on which quantities are estimated,

two theoretical approaches have been taken.

One is to continue with algebraic models, modifying the ordering relationship. It is postulated that apparent error in choices arise from trying to fit a nontransitive indifference relation by a transitive one. This presupposes that strict inequality is actually transitive and places all the difficulty in the indifference relation.

Two classes of relations, called interval and semiorders, which correspond in a reasonable way to the existence of thresholds have been studied. A summary will be given in Chapter 15 of volume Two of Foundations of Measurement. The reader is also referred to survey articles by P. C. Fishburn in Operations Research, and by Roberts in the Journal of Mathematical Psychology.

The other approach seems somewhat more promising to the lecturer. Here we assume that the algebraic models are idealizationsof underlying probabilistic models, and so the attempt should be to capture this structure. Thus, we assume that people make choices not according to the strict preference, but probabilistically, i.e., not $a \geq b$, rather $Pr(a \geq b)$. In the special binary case where $P = \{0, \frac{1}{2}, 1\}$ the model should reduce to the algebraic formulation although as yet that goal is not really fulfilled. The problem seems to be that there is no very natural way of putting together the underlying algebraic structure of either a semigroup operation or a Cartesian product with the probabilistic ordering -- more of that later.

Some references of interest for the following discussion include Luce and Suppes (1965), Tversky (1972) and Tversky and Russo (1969).*

Probabilistic Ordinal Theory

Let the set (finite or infinite) of objects among which a person has preference be denoted A (for alternatives). Any finite subset, R, S, ... may be presented from which the subject is to choose one. For x ε R, denote by $p_R(x)$ the probability that he chooses x when R is presented. Put another way, when R is the choice set, it is taken to be a sample space and it has a probability measure p_R. Note that this is not a conditional probability; in particular for R $\subset$ S, $p_R(x)$ need not equal $p_S(x|R)$. Obviously, however we do anticipate some sort of relation among the several probability measures, and one task facing psychologists is to try to discover it.

Most of the proposals that have been made have a common theme, namely, the existence of an underlying numerical quantity, usually called utility, from which the probability

---

*Luce, R.D. and P. Suppes (1965) "Preference, Utility and Subjective Probability," in R.D. Luce, R.R. Bush and E. Galanter (eds.) Handbook of Mathematical Psychology, Vol. 3, New York, Wiley, pp. 249-410 (see 331-350 in particular).

Tversky, A. (1972) "Elimination by Aspects: A Theory of Choice", Psychological Review, 79:281-299.

Tversky, A., and J. E. Russo (1969) "Substitutibility and Similarity in Binary Choices", Journal of Mathematical Psychology 6:1-12.

is somehow computed. Explicitly, let us suppose that there
is a function $u: A \to Re$ and functions F_n of n arguments such
that when $|R| = n$,

$$p_R(x) = F_n[u(x); u(y_1), \ldots, u(y_{n-1})] .$$

Perhaps the best known example of this type is the one called
the strict utility model, in which the F's are of the form

$$p_R(x) = \frac{u(x)}{\sum\limits_{y \in R} u(y)} \quad , \ y \in R \subset A .$$

The lecturer's work on this model is reported in Luce (1959).*

For the present, however, let us examine the general
idea, but only for choice between pairs. For simplicity of
notation, we write

$$p(x,y) \equiv p_{\{x,y\}}(x) = F[u(x), u(y)] .$$

Intuitively, if we either increase the utility of x or de-
crease that of y, we anticipate an increase in $p(x,y)$, which
means we must assume F is strictly decreasing in the second
one. When the binary choice probabilities are of this form,
they are said to satisfy simple scalability.

Although simple scalability does not seem very restrictive,
it is an empirical question whether preference (or other)
judgments fulfill it. One problem in answering this, is to
transform the theoretical notion of simple scalability into

---

*Luce, R. D. (1959) Individual Choice Behavior: A
Theoretical Analysis, New York, Wiley.

is somehow computed. Explicitly, let us suppose that there is a function $u:A \rightarrow Re$ and functions F_n of n arguments such that when $|R| = n$,

$$p_R(x) = F_n[u(x); u(y_1),\ldots, u(y_{n-1})] \ .$$

Perhaps the best known example of this type is the one called the strict utility model, in which the F's are of the form

$$p_R(x) = \frac{u(x)}{\sum\limits_{y \epsilon R} u(y)} \quad , \ y \ \epsilon \ R \subset A \ .$$

The lecturer's work on this model is reported in Luce (1959).*

For the present, however, let us examine the general idea, but only for choice between pairs. For simplicity of notation, we write

$$p(x,y) \equiv p_{\{x,y\}}(x) = F[u(x),u(y)] .$$

Intuitively, if we either increase the utility of x or decrease that of y, we anticipate an increase in p(x,y), which means we must assume F is strictly decreasing in the second one. When the binary choice probabilities are of this form, they are said to satisfy simple scalability.

Although simple scalability does not seem very restrictive, it is an empirical question whether preference (or other) judgments fulfill it. One problem in answering this, is to transform the theoretical notion of simple scalability into

---

*Luce, R. D. (1959) Individual Choice Behavior: A Theoretical Analysis, New York, Wiley.

a testable property, i.e., one that is stated wholly in
terms of the probabilities, without mention of either u
or F. The following formulates three observable properties
that are each equivalent to simple scalability.

PROPOSITION 1. Assuming binary probabilities not equal to
0 or 1, the property of simple scalability is equivalent to
each of the following:

Strict stochastic transitivity (SST): If $p(x,y) \geq 1/2$ and
$p(y,z) \geq 1/2$, then $p(x,z) \geq \max [p(x,y),p(y,z)]$;moreover if both
$\geq$ in the hypothesis are $>$, then the conclusion is $>$.

Substitutability: $p(x,z) \geq p(y,z)$ iff $p(x,y) \geq 1/2$.

Independence: $p(x,z) \geq p(y,z)$ iff $p(x,w) \geq p(y,w)$.

Proof: Suppose p satisfies simple scalability and that
$p(x,y) \geq 1/2$ and $p(y,z) \geq 1/2$. Since F is strictly increasing
in the first argument and

$$p(x,y) = F[u(x),u(y)] \geq 1/2 = F[u(y)] = p(y,y),$$

$u(x) \geq u(y)$. And so

$$p(x,y) = F[u(x),u(y)] \geq F[u(y),u(z)] = p(y,z).$$

The other inequality follows similarly from the fact that F is
decreasing in the second argument. The case of strict in-
equalities follows from the strict monotonicity of F.

Suppose SST holds, but that substitutability is false,
i.e., either

(i) $p(x,z) > p(y,z)$ and $p(y,x) \geq 1/2$

or (2) $p(x,z) = p(y,z)$ and $p(x,y) \neq 1/2$.

For (i), there are two subcases. If $p(x,z) \geq 1/2$, then SST

implies $p(y,z) \geq p(x,z)$, contrary to hypothesis. If
$p(y,z) < 1/2$, then $p(z,y) > 1/2$ which with $p(y,x) \geq 1/2$
yields (SST) $p(z,x) \geq p(z,y)$, contrary to hypothesis. For
(ii) the arguments are similar and are left to the reader.

Assuming substitutability, independence is an immediate
consequence.

Finally, we assume independence and prove simple
scalability. Fix z, and set $u(x) = p(x,z)$. Define F by

$$F[u(x),u(y)] = p(x,y).$$

F is well defined for if $u(x) = u(y')$, then $p(x,z) = p(x',z)$,
whence by independence, $p(x,y) = p(x',y)$. Also, $p(y,z) =$
$p(y',x)$, and so $p(y,x') = p(y',x')$. Therefore,

$$F[u(x),u(y)] = p(x,y) = p(x',y') = F[u(x'),y(y')].$$

Reversing this argument shows that F is 1:1 in each argument.
To show it is strictly increasing in the first one, suppose
$u(x') > u(x)$, i.e., by definition, $p(x',z) > p(x,z)$, whence
by independence

$$F[u(x),u(y)] = p(x',y) > p(x,y) = F[u(x),u(y)].$$

The proof that F is strictly decreasing in the second com-
ponent is similar.

The following example was first given by Savage to show
that the strict utility model cannot be generally correct; it
applies to any case of simple scalability. Suppose a young
child is presented with a choice between a bicycle (x), and
a pony (z). He vacillates at the bicycle dealer's, and to
enhance the attractiveness of the bicycle, the dealer offers a
siren on the bicycle (y). Evidently $p(y,x) \gg \frac{1}{2}$. Yet

clearly, if the child wanted the bicycle in the first place
he would still want the bicycle, and if he wanted the pony
the additional utility of the siren would be most unlikely
to change his mind. So we conclude $p(x,z) = p(y,z)$. But
these two probability statements are inconsistent with
substitutibility, and so with the hypothesis of simple
scalability.

Beyond this experiment simple scalability is not satis-
fied in actual experimental data. In preferences for greys,
C. H. Coombs (J. exp. Psychol., 1958, 55, 1-7) has shown that
certain selected triples of stimuli -- chosen so that they
span the subject's ideal grey -- systematically violate SST.
D. H. Krantz (J. math. Psychol, 1967, 4, 226-245) has shown
that simple scalability does not hold for judgments of
similarity between pairs of monochromatic colors. Specifically,
substitutability is violated since whether or not one alterna-
tive can be substituted for another depends upon the context.
Finally, Tversky and Russo (J. math. Psychol, 1969, 6, 1-12)
have exhibited serious failures in independence when subjects
judge the size of rectangles and lenses of various proportions.
"It was found that the similarity between stimuli facilitates
the discrimination between them. But since the similarity
between two stimuli can be varied without changing their
scale values, simple scalability, and hence independence, must
be violated." (p. 11).

These data require that we look for a more general class
of models than those just considered. One important proposal

retains the original algebraic idea of a numerical utility
function over the alternatives and that, of the alternatives
presented, the one having the greatest utility is selected.
The new wrinkle is that we do not assume that the utility
stays put each time a choice is made. Rather the utility
associated with each alternative is a random variable, and
the alternative with the greatest existing utility at the
time of choice is selected. In particular, let $U(x)$ be
the random variable associated with alternative x (the sample
space is implicit), then the probability of selecting x from
R is

$$p_R(x) = Pr[U(x) \geq U(y), \text{ for all } y \in R].$$

In this case, the choice probabilities are said to satisfy a
random utility model, which is an independent one if the
random variables are independent.

Some questions must be asked about the model. First, is
it really more general than simple scalability? The answer
is yes. Second, are there examples of simple scalability that
are also random utility models? The answer is yes, and one
case is formulated as:

PROPOSITION 2: If a set of choice probabilities satisfy a strict
utility model, then they satisfy an independent random utility one
Proof. Let the v's defining the strict utility model be given,
and let $U(x)$, $x \in A$, be independent random variables with distri-
bution functions

$$Pr\left[U(x) \leq t\right] = \begin{cases} e^{v(x)t}, & t \leq 0 \\ 1, & t \geq 0 \end{cases}$$

Now

$$\Pr\left[U(x) \geq U(y), y\varepsilon R\right] = \int_{-\infty}^{\infty} \Pr\left[U(x) = t\right] \prod_{y\varepsilon R-\{x\}} \Pr\left[U(y) \leq t\right] dt$$

$$= \int_{-\infty}^{0} v(x) e^{v(x)t} \prod_{y\varepsilon R-[x]} e^{v(y)t} dt$$

$$= \int_{-\infty}^{0} v(x) \exp\left(\Sigma_{y\varepsilon R} v(y)t\right) dt$$

$$= \frac{v(x)}{\sum_{y\varepsilon R} v(y)}$$

$$= p_R(x) \ .$$

And, third, what properties of the probabilities are equivalent to this model, and do the data satisfy these properties? No general answer is now known to this question. However, one necessary property is known. This says that any reduction in the size of the choice set never decreases the probability of choosing one of the remaining alternatives. Formally, the choice probabilities are _regular_ if for all $x \ \varepsilon \ R \subset S$, then

$$p_R(x) \geq p_S(x).$$

An open problem is to find observable properties that are equivalent to the random utility model.

Some other references to the work in constant utility models and random variable models are to be found in Luce and

Suppes (1965).

More Recent Developments

Tversky (1972) has constructed a model which overcomes the difficulties in the constant utility model, and which, intuitively, makes some psychological sense. He calls his model the Elimination by Aspects (EBA) model and motivates it as follows (Tversky, 1972:296):

"The EBA model accounts for choice in terms of a covert elimination process based on sequential selection of aspects. Any such sequence of aspects can be regarded as a particular state of mind which leads to a unique choice. . . . According to the present theory, choice probability is an increasing function of the values of the relevant aspects. Indeed, the elimination-by-aspects model is compensatory in nature despite the fact that at any given instant in time, the choice is assmued to follow a conjunctive (or lexicographic) strategy.

In the proposed model, aspects are interpreted as desirable features; the selection of any particular aspect leads to elimination of all alternatives that do not contain the selected aspect."

The model will be formulated with Tversky's notation, which differs only in minor ways from the notation that has been used up to this point.

Let T be the set of alternatives, and T' be the corresponding set of aspects.

Define a function $f: T \rightarrow 2^{T'}$, i.e., a mapping from T to the power set of T'. So for each $x \in T$, $f(x)$ is interpreted

as the set of aspects associated with x, and is abbreviated
as x'. For each $A \subset T$, define $A' = \{\alpha | \alpha \in x'$ for some
$x \in A\}$ and $A^0 = \{\alpha | \alpha \in x'$ for all $x \in A\}$. If $\alpha \in T'$, let
$A_\alpha = \{x | x \in A \ \& \ \alpha \in x'\}$.

The model postulates that there exists a positive weighting
function u on $T' - T^0$ (u: $T' - T^0 \to Re^+$), such that

$$\frac{\sum\limits_{\alpha \in A'-A^0} u(a)p_{A\alpha}(x)}{\sum\limits_{\beta \in A'-A^0} u(\beta)}$$

This is a recursive expression in the choice probabilities.
Several properties of the model are:

1. If the alternatives are such that they have no aspects in
common, then the model reduces to the strict utility model,
i.e., the expression above takes the form

$$\frac{v(x)}{\sum\limits_{y \in A} v(y)}$$

2. The model satisfied <u>regularity</u>, i.e.,

$$x \in A \subset B \to p_A(x) \geq p_B(x)$$

3. The model exhibits moderate stochastic transitivity, i.e.,

$$p(x,y) \geq \tfrac{1}{2} p(y,z) \geq \tfrac{1}{2} \to p(x,z) \geq Min[p(x,y) \ p(y,z)] \ .$$

Although data have been reported in which strict stochastic
transitivity fails, it appears that moderate stochastic transi-
tivity is sustained. Of course, these statements must be
accepted with some caution because we never have probabilities

themselves, but only estimates of them. Thus if
$p(x,y) \geq \frac{1}{2}$ is near $\frac{1}{2}$, it is perfectly possible for
$\hat{p}(x,y) < \frac{1}{2}$ etc.. Thus, even if a property such as moderate
stochastic transitivity were strictly true, apparent violations
of it may very well occur in the estimates. Since the
statistical decision problem has never been properly formulated
for these kinds of statements, it is a matter of judgment
whether the model is invalid or not.

Probabilistic Additive Theory

Since we know little about the probabilistic version of
ordinal choice, it is not surprising that we know even less
about how to combine the probability model with the additive
structure of extensive or of conjoint measurement. Three
minor observations can be made:

1. Physicists make the assumption that measurement errors
have a normal distribution in extensive measurement. This
cannot be strictly correct since a normal variable can have
arbitrarily large negative values whereas the measurements
themselves are always positive. What we would like is a
theory of the observed random variables. The theory should
exhibit two properties, at least. First, the random
variables should be positive. Second, one should not be
able to take advantage of the decomposability of extension
measures to reduce the error of measurement below that given
by the distribution function of the theory.

If we let $\Phi(a)$ denote the random variable associated with

object a, then the latter property suggests,

$$\Phi(aob) = \Phi(a) + \Phi(b)$$

from which it follows that

$$E\Phi(aob) = E\Phi(a) + E\Phi(b),$$

and so $E\Phi$ behaves like the ordering extension measure. Assuming independence, the distribution function of $\Phi(aob)$ will be the convolution of that of $\Phi(a)$ with $\Phi(b)$. It seems plausible to demand that the family of distributions should be closed under convolutions. One well known family that has this property is the gamma,

$$P[\Phi(a) = t] = \frac{\lambda^{\phi(a)} t^{\phi(a)-1} e^{-\lambda t}}{\Gamma[\phi(a)]}$$

In this case,

$$E\Phi(a) = \frac{\phi(a)}{\lambda} ,$$

so we must assume

$$\phi(aob) = \phi(a) + \phi(b)$$

The family also approaches the normal distribution as $\phi(a)$ becomes large. A problem, then, is to give a plausible axiomatization of probabilistic extension measurement which leads to this (or some other similar) distribution function.

In the conjoint measurement case we have the problem of finding a reasonable axiomatic structure that will handle the natural discontinuities that must occur. Suppose $A_1, \geq_1$ and $A_2, \geq_2$ is an ordering (say, by preference) then surely $p(ap \geq bq) = 1$, when $a \geq_1 b$, $p \geq_2 q$. Yet a small

reduction sufficient to reverse the first inequality may cause a large change in p, say to near $\frac{1}{2}$.

The Measurement of Color

The material presented in this lecture is part of Chapter 14 of Volume Two of Krantz et al., which is in preparation.

The physical description of incoherent light (omitting spatial and temporal variations) is given by the radiant energy density at each wave-length in the visible part of the electromagnetic spectrum. Thus we have functions $a(\lambda)$ defined for between red beginning at 400 nanometers (nm) through violet ending at 700 nm (1 nm = 10^{-9} meters), giving the radiant energy density at wave-length λ.

One of the intriguing aspects of color perception has to do with the range of possible combinations of colored light. For example, one is familiar with such combination as greenish-blue, and reddish yellow, but a reddish green is never perceived, nor is bluish yellow. A combination of red and green will appear either as grey or red or green, but not a combination. (The colors will be attentuated, but never mixed.) We mention this only as evidence that there is more to the perception of color than simple superposition of the percepts due to monochromatic lights.

Similarly, it is of interest that colors that are visually matched may be composed of quite distinct frequency distributions of wave-lengths. These two characteristics do

not hold for auditory stimuli. It does not occur that sub-
jects match sounds which are generated by different distri-
butions of energy over auditory frequency, and so auditory
stimuli do not exhibit the same sort of compression of in-
formation.

A Theory of Color

The following primitive notions are utilized in develop-
ing a theory of color. We begin with a set A of all spectral
distributions of $a(\lambda)$ and two physical operations, $\oplus$: A X A $\rightarrow$ A,
which represent the addition of the distributions $a(\lambda) + b(\lambda)$,
and *: Re^+ X A $\rightarrow$ A, which permits us to change all energies
by a given factor, i.e., $ta(\lambda)$. The idealized psychophysical
data take the form of an equivalence relation $\sim \subset$ A X A where
$a(\lambda) \sim b(\lambda)$ represents the judgment that the two distributions
$a(\lambda)$ and $b(\lambda)$ are of the same color.

Definitions

The structure <A, $\oplus$, *> is a convex cone if

      (i) $\oplus$ is associative, commutative and satisfies
            the cancellation property, i.e., if $a \oplus c = b \oplus c$
            then $a = b$.

     (ii) $t*(u*a) = (t*u)*a$

            $t*(a \oplus b) = (t*a) \oplus (t*b)$

            $(t+u)*a = t*a \oplus u*c$

            $1*a \quad\quad = a$

Behavioral Aspects

The system $\langle A, \oplus, *, \sim \rangle$ is a Grassman structure (after the mid 19th century theorist who worked on color and stated the properties of the equivalence relation). It has the following properties.

(1) $\langle A, \oplus, * \rangle$ is a convex cone

(ii) $\sim$ is an equivalence relation (with the usual problems of assuming an empirical relation is an equivalence relation).

(iii) $a \sim b$ iff $a \oplus c \sim b \oplus c$, (i.e., if two colors match and one adds (subtracts) the same light to (from) both, they will continue to match).

(iv) $a \sim b$ implies $t*a \sim t*b$. (This is valid for a large range of distributions. It does not hold for very low intensities.)

Definition of m-chromatic

A structure is m-chromatic iff. there exists an m-element basis $a_1, \ldots, a_m$ such that $\sum_1^m t_i * a_i \sim \sum_1^m u_i * a_i \Rightarrow t_i = u_i$. A basis is maximal in the sense that there exists $a_0, \ldots, a_m$ and $t_i \neq u_i$, such that $\sum_0^n t_i * a_i \sim \sum_0^m u_i * a_i$. (Only m of these are basic and they can generate the rest.)

The Trichromatic Nature of Color

There is very good evidence (contrary to a common mis-interpretation of the Land experiments) that color is 3-chromatic in the following sense, given the three suitably

chosen colors A_1, A_2, A_3 (for example, red, green and blue) it is possible to take any distribution a, and find coefficients t_i such that either

(i) $a \sim \sum\limits_{i=1}^{3} t_i {}^* a_i$

or   (ii) $a \oplus t_1 a_1 \sim t_2 {}^* a_2 \oplus t_3 {}^* a_3$

or   (iii) $a \oplus t_2 {}^* a_2 \sim t_1 {}^* a_1 \oplus t_3 {}^* a_3$

or   (iv) $a \oplus t_3 {}^* a_3 \sim t_1 {}^* a_1 \oplus t_2 {}^* a_2$

is empirically true.

Theorem (see Chapter 14, in Vol. II of Krantz <u>et</u>. <u>al</u>. for proofs)

If <A, $\oplus$, $*$> is a Grassman structure, there exists a vector space <V, +, .> over the reals and a convex cone $C \subset V$, and a function ϕ: A $\rightarrow$ C, such that

(i) $\phi(a \oplus b) = \phi(a) + \phi(b)$

(ii) $\phi(t {}^* a) = t\phi(a)$

(iii)   if $x \varepsilon V$, there exist c,d such that $x = \phi(c) - \phi(d)$

(iv) $a \sim b$ iff $\phi(a) = \phi(b)$

Then if we have another function with the same structure they are all related by a non-singular linear transformation.

Corollary

If the Grassman space is m-chromatic, then the vector space is m-dimensional.

In practice, the real vector space is utilized using the above matching data and assigning

$$\phi(a) = \begin{cases} t_1, t_2, t_3 & \text{if} \quad \text{(i)} \\ -t_1, t_2, t_3 & \text{if} \quad \text{(ii)} \\ t_1, -t_2, t_3 & \text{if} \quad \text{(iii)} \\ t_1, t_2, -t_3 & \text{if} \quad \text{(iv)} \ . \end{cases}$$

There are various standardized coordinate systems. The choice of a particular coordinate systems has been made on at least one of three criteria: ease of standardization, computational convenience, or representation of additional empirical relations on colors, beyond $\oplus$, *, and $\sim$. These criteria have led to three different standard coordinate systems adopted by the CIE (Commission Internationale d' Éclairage). One of these,the R,G,B system, is based on red, green and blue instrumental primaries that can easily be standardized. Another, the X,Y,Z system, corresponds to no possible set of instrumental primaries; rather the matrix (α_{ij}) is chosen such that all the coordinates ϕ_2' are non-negative, and such that the second (ϕ_2' or y) coordinate represents luminance (approximately, another empirical rela-tion on colors based on certain special methods of brightness matching). This system is extremely convenient computationally, because of positive numbers and because the luminance of light is directly represented by one coordinate. The third set of coordinates is the uniform chromaticity system (UCS), which attempts to give an approximate representation of empirical color-discriminability and color-similarity relations.

by means of Euclidean distances that take the UCS ϕ' coordinates as orthogonal axes.

A standard way of representing chromaticity coordinates can be seen in the following figure which gives the chromaticity diagram for the X,Y,Z coordinate system of the CIE. The abscissa is $\bar{\phi}_1 = x = X/(X + Y + Z)$; the ordinate is $\bar{\phi}_2 = y = Y/(X + Y + Z)$. The curved locus of monochromatic lights (spectrum locus) is shown, together with the locus of a number of other lights. Point A is the locus of the light distribution from a tungsten filament lamp (maetameric to black-body radiation at 2842°K), and the other points shown by two-digit numbers that correspond to that light distribution after modification by Kodak Wratten gelatin filters with the corresponding numbers.

Theoretical Elaborations of Color Measurement.

The theoretical elaboration that shall be examined here has to do with coding. Investigators attempt to discover the nature of perceptual processes under the assumption that sense data has to be coded information, and codes are defined as numerical functions operating on the stimuli. A potential code appears to be derivable from the fact that different pigments in the eye absorb light quanta differentially. The fact that such pigments exist seems clear -- the slow recovery of the eye to stimulation indicates that pigments are bleached and reinstated over time. There are a number of different pigments, and it seems logical to utilize the
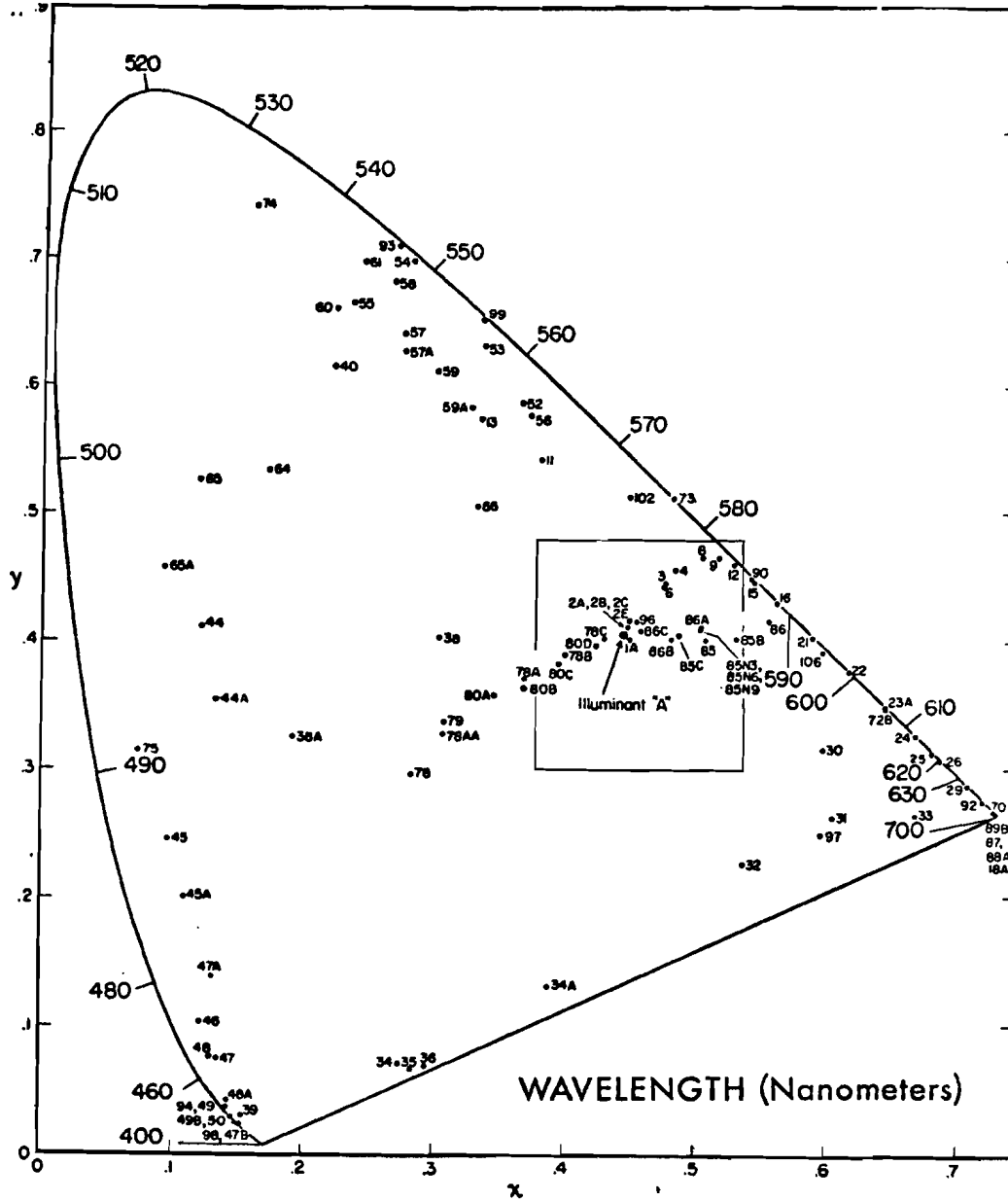
FIGURE 2. Chromaticity coordinates, showing the spectrum locus
and loci of lights from a tungsten lamp (CIE Standard Source "A")
modified by various Kodak gelatin filters. (From Kodak Filters
for Scientific and Technical Uses. Eastman Kodak Co. B-3, 1970.)

proportion of quanta captured by the pigment to devise
quantum-catch functions of the following kind

$$p_j(a), \quad j = 1,\ldots,\kappa.$$

There are most likely three photopigments and they are complete in the sense that a function $g:A \to Re$ is a code if
$a \sim b$ then $g(a) = g(b)$. A set g_i of codes is <u>complete</u> if
$g_i(a) = g_i(b)$, $i = 1,\ldots,n$, iff $a \sim b$. A code is <u>linear</u>
if $g(a \oplus b) = g(a) + g(b)$ and $g(t*a) = tg(a)$.

But it is not yet clear exactly how the pigments relate
to the codes that are suggested by various psychological
phenomena.

For example, let us examine the psychological theory
first proposed by Hering in the late 19th century and
developed in the past twenty years by Hurvich and Jameson of
the University of Pennsylvania. The basic idea is as follows.
For any stimulus we attempt to ascertain the amount of redness or greenness and the amount of yellowness or blueness
it exhibits. Select a_1 to be a greenish color and b_1 to be
a reddish color, such that $a_1 \oplus b_1$ is neither. The empirical
question if a is reddish rather than greenish is how much a_1
has to be combined with a to get rid of the reddishness, i.e.,
find t such that $t*a_1 \oplus a$ is neither red nor green. If a is
reddish, then we find t such that $t*a_1 \oplus a$ is neither red nor
green. Similarly we determine relative to a given yellow
and glue, how much yellow or blue a contains. For monochromatic hues we may plot the value of t versus wave length to

show the nature of color perception as a function of wave length.

Figure 3

Color Perception with Changing Wave Length
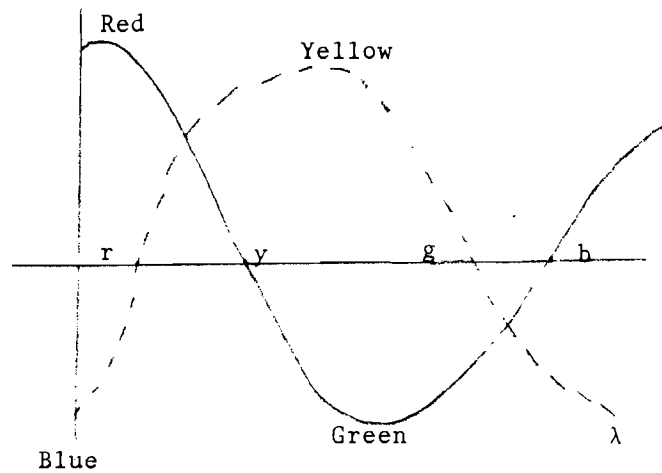
Note that at wave length y the stimulus has neither red nor blue, and so it is pure yellow; at point g it is pure green, at r red; and at b blue. Empirically the points r and y are stable across subjects, whereas g and b are not so stable, leading to different perceptions as to the relative amounts of green and blue in a stimulus.

To formulate this idea as a theorem, we need several definitions. In a Grassman structure <A, ⊕, *, ∿>, define

$$En = \{a \mid a⊕b∿b\}$$

The structure is <u>proper</u> if En = φ.

If E ⊂ A, then <A, ⊕, *, E> is an equilibrium structure if

(i) $a \in A$ implies there exists $a' \in A$ such that $a \oplus a' \in E$.

(ii) Suppose $a \in E$. Then $b \in E$ if $a \oplus b \in E$.

(iii) $a \in E$ implies $t^* a \in E$.

In such a structure we may define $\sim_E$ as

$a \sim_E b$ iff there exists C such that $a \oplus c \in E$ iff $b \oplus c \in E$.

It can be shown that if $<A, \oplus, *, E>$ is an equilibrium structure, then $<A, \oplus, *, \sim_E>$ is a Grassman structure, and also we can speak of the dimensionality of the equilibrium structure as that of its individual Grassman structure.

In the following result, think of A_1 as the set of stimuli that are neither reddish nor greenish and A_2 as those that are neither greenish nor blueish.

Theorem. Suppose $<A, \oplus, *, \sim>$ is a proper trichromatic Grassman structure with $A_i \subset A$, $i = 1,2$. Assume,

(1) (Code axiom) $a \sim b$ and $a \in A_i$ implies $b \in A_i$

(2) (Linearity) If $a \in A_i$ then $t^* a \in A_i$ and $b \in A_i$ iff $a \oplus b \in A_i$.

(3) (Hue components) There exist $a_i, b_i \in (A_1 \cap A_2) - A_i$ such that $a_i \oplus b_i \in A_i$, then

    (i) $<A, \oplus, *, A_i>$ is a 1-dimensional equilibrium structure

    (ii) $<A, \oplus, *, A_1 \cap A_2>$ is a 2-dimensional equilibrium structure

and $<A_1 \cap A_2, \oplus, *, \sim>$ is a proper 1-chromatic Grassman structure.

    (iii) There exist real ϕ_i, $i = 1,2,3$, such that

(ϕ_1, ϕ_2, ϕ_3) represents $<A, \oplus, *, \sim>$

(ϕ_1, ϕ_2) represents $<A, \oplus, *, A_1 \cap A_2>$,

ϕ_1 represents $<A, \oplus, *, A_i>$, $i = 1,2$ and

ϕ_3 represents $<A_1 \cap A_2, \oplus, *, \sim>$. This is unique up to

$$\begin{pmatrix} \alpha_1 & 0 & 0 \\ 0 & \alpha_2 & 0 \\ \beta_1 & \beta_2 & \beta_3 \end{pmatrix}$$

(iv) If, in addition, there exists $\sim_b \supset \sim$ (inter-
preted as equal brightness) such that
$<A, \oplus, *, \sim_B>$ is a proper 1-chromatic Grass-
mann structure, then $\sim_B = \sim / A_1 \cap A_2$ and ϕ_3
can be chosen to represent $\sim_B$ and the unique-
ness is then,

$$\begin{pmatrix} \alpha_1 & 0 & 0 \\ 0 & \alpha_2 & 0 \\ 0 & 0 & \alpha_3 \end{pmatrix}$$

when α_1, $\alpha_2 \neq 0$ and $\alpha_3 > 0$.

Currently work is being carried out attempting to relate
this psychological code to the photopigments. This is subject
to various empirical constraints, such as providing a natural
account of the several kinds of color blindness (see Chapter 14).

The Psychophysics of Auditory Intensity

In the final section of these lectures we will explore
some experimental and theoretical work concerned with
various aspects of the perception of changes in auditory
intensity. Some of the results may apply to intensive as-
pects of other stimuli, such as brightness, but there are
so many differences when one passes from the study of one
modality to another that one cannot count on a direct carry-
over. We begin by giving a sketchy background of some
psychophysical methods.

The basic questions that we explore have to do with how
the organism manages to process discrete changes in auditory
stimuli and to answer simple questions about these changes.
These questions are whether a change in auditory stimuli has
occurred (detection), which of several possible changes has
occurred (recognition), which of two changes is the larger
(discrimination), and how large is a given change (scaling).

In the classical view, the organism is perceived as a
machine, or a transducer, operating on a signal. Even though
experimenters knew this wasn't correct -- for example subjects
make errors and the error rates were maintained at low levels
(two to five per cent) by stern reprimands -- they believed
that anything aside from the sensory transduction was ex-
traneous and relatively unimportant. But during World War
II, with the growth of interest in sonar and radar signal
detection, a theory of ideal signal detection was worked out

and by the early 1950's psychologists had begun to adapt this theory of signal detectibility to their own use (see Green and Swets 1966)*. One of the more important empirical phenomena to arise from this work is the ROC curve ("receiver operating characteristic curve") or iso-sensitivity curve** which can be described as follows.

Imagine a situation in which one of two possible signals, s_0 or s_1 is presented on each trial and the observer tries to identify which has been presented. One of the signals could be the null signal, in which case it is a detection experiment; otherwise it is a recognition study. Suppose the choice of presentations is random, with s_1 occurring with probability P. Associated with every signal-response pair is a payoff which also serves as information feedback. We can summarize the experimental information in the following matrix:

$$
\begin{array}{cc}
 & \begin{array}{cc} r_1 & r_0 \end{array} \\
\begin{array}{cc} P & s_1 \\ 1-P & s_0 \end{array} & \begin{bmatrix} o_{11} & o_{10} \\ o_{01} & o_{00} \end{bmatrix}
\end{array}
$$

Assuming that the responses are statistically independent of both the signal presentations and the response on earlier

---

*Green, D. M., and J. A. Swets (1966) Signal Detection Theory and Psychophysics. New York. Wiley.

**"Iso-sensitivity" is an alternative term proposed in Luce, R. D. (1963) "Dectection and Recognition" in R. D. Luce, R. R. Bush, E. Galanter (eds.) Handbook of Mathematical Psychology. New York. Wiley, pp. 103-189.

trials, as is approximately the case for well-trained observers, the observed frequencies N_{ij} of response r_j to the presentation of signal s_i can reasonably be thought of as providing us with estimates of underlying conditional response probabilities. In particular, if we abbreviate

$$p_{ij} = \Pr(r_j | s_i)$$

where

$$p_{i1} + p_{i2} = 1, \; i = 0,1$$

then we estimate p_{ij} by

$$\hat{p}_{ij} = \frac{N_{ij}}{N_{i0} + N_{i1}} \; .$$

Observe that there are only two independent conditional probabilities, and so the data are summarized by just two numbers. It is conventional in this literature to use p_{01} and p_{11}, whereas statisticians usually focus on the two error p_{01} and p_{10} .

The important empirical fact is that the signals alone do not uniquely determine p_{01} and p_{11}, as was implicit in the classical view. Rather, if we vary the presentation probability P or the payoff 0_{ij} or the instructions to the observer, under each motivational condition we obtain a different pair of numbers. Moreover, the values are much too different to be accounted for by variability in the estimates. Indeed, they appear to run from (0,0) to (1,1) and to arise from a smooth
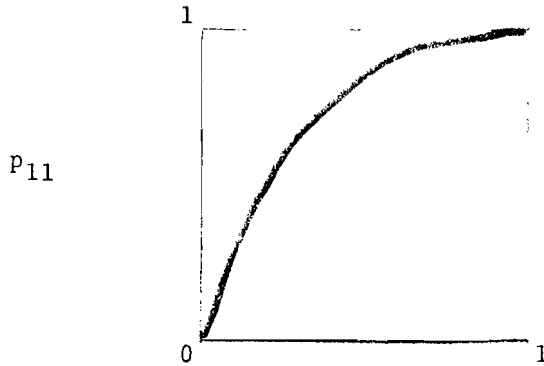
curve such as that shown in Figure 4.

Figure 4. The ROC Curve

The classical approach focussed attention on p_{11} and ignored p_{01} except to keep it small. Note that in this region the slope is large and so the value of p_{11} is very sensitive to small errors in estimating p_{01}.

One curious example of ROC analysis is A. Stunkard's (University of Pennsylvania) data on obesity. In this work he placed a baloon in the stomach of his subjects in order to study the relationship between stomach contractions and re-ported hunger, Let s_1 denote a time period with a contraction and s_0 one with no contractions, and let response 1 denote hunger, and 0 no hunger, then each subject is represented by a point in the unit square. Stunkard found that the data from a number of subjects appeared to sweep out an ROC curve, but that there were variations in the reports of hunger: obese men overreported, and obese women underreported in comparison to normals, but all seemed about equally sensitive to contractions.

There is another way of plotting ROC curves. If p is
a probability and N (0,1) denotes the normal density with
mean 0 and variance 1, then the normal deviate is the value
Z such that

$$p = \int_{-\infty}^{z} N (0,1).$$ Plotting ROC data in

this manner usually yields an approximately straight line
representation of the following kind:
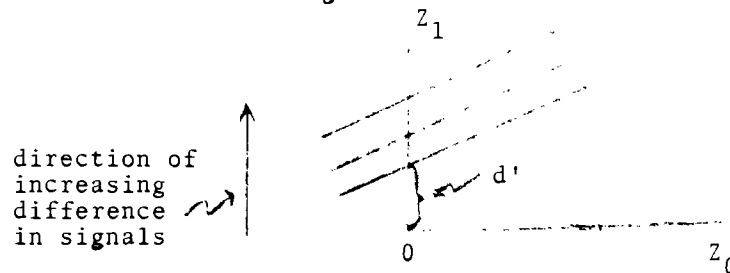
indicating that the underlying distributions are approximately
normal. An increase in the intensity separation of signals
generates a family of more-or-less parallel lines that move
upward. The value of Z_1, corresponding to $Z_0 = 0$ is a measure
of detectability; it is denoted by d'.

Another approach to signal detection involves two inter-
vals with one signal in the first and the other signal in
the second interval. So the stimuli may be written $<s_0, s_1>$
and $<s_1, s_0>$. The subject is required to report which interval
contains the more intense signal. The subject can be expected
to do better in this design than in the single interval one,
since he has two trials, as it were. We attempt to see how
this two-alternative forced choice procedure relates to the
single interval case.

Let us suppose that when signal s_i is presented it re-
ceives an internal representation X, which is a random
variable with density f_i (which is approximately normal).
His response rule is to establish a criterion Z and to say
all observations $X < Z$ are called signal 0, and all $X \geq Z$
are 1. Thus,

$$P_{i1}(Z) = \int_Z^\infty f_i(x)dx, \quad i = 0,1.$$

Graphically the situation is as follows:
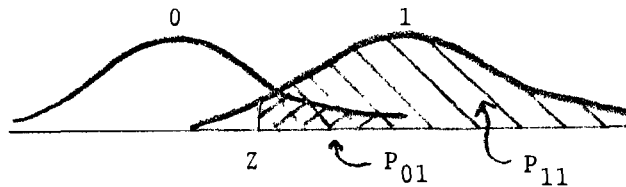
Turning to the two interval case, let us consider the proba-
bility of a correct response on the assumption that the two
observations X_1 and X_2 are independent random variables. We
suppose that he reports s_1 to be located in the interval with
the larger observation. The probability of a correct response
is simply the probability that the larger observation occurs
with s_{1i}, i.e.,

$$p(C) = \int_{-\infty}^\infty f_i(x) \int_{-\infty}^x f_i(y)dy.$$

Making the change of variables from above,

$$1 - p_{21}(x) = \int_{-\infty}^{x} f_2(y)dy$$

$$- dp_{11}(x) = f_1(x) \ ,$$

$$p(C) = \int_{0}^{1} [1 - p_{21}(x)] \ dp_{11}(x)$$

which is easily seen to be the area over the one interval.
In this way a connection is forged between the two kinds of
experiments. Note that we make the assumptions that the
subject is operating on an underlying random variable whose
distribution is uniquely determined by the signal presentation
and that the decision procedure operates on this representa-
tion of the signal.

In the work about to be described, two features distin-
guish the approach from both the classical and signal de-
tectability approach. First, time is included in the process
description -- and second neurophysiological data that have
been collected on single auditory fibres during the past ten
years is used to suggest the nature of the model.

Physiological Data

An auditory input can be characterized as a continuous
function of time, for example, as pressure at the eardrum
as a function of time. When one measures the electrical
activity in individual nerve fibres of the peripheral auditory
nervous system, one does not see any direct analogue of that

function. Rather, each fibre conducts a train of electrical
impulses, each of which is of very brief duration (a few
microseconds) and is of approximately the same voltage. At
first sight, these pulse trains are highly irregular: some-
times they are obviously affected by changes in the signal,
at other times apparently not.

For a while it was felt that the impulse rate was important
primarily in determining which fibres are active, and the
information is encoded in the pattern of active fibres. These
"place theories" of audition, while plausible in neurophysio-
logical terms, had difficulty in accounting well for the psy-
chophysical data. In particular it has been difficult to re-
late sensibly the results on frequency and intensity discrimina-
tion. During the 1940 to 1960's, first Galambos and Davis
(1943)* then Kiang (1965, 1968) and Rose and his associates
(1967) were able to implant microelectrodes into single nerve
fibres and measure the temporal pattern of impulses with an

---

*Galambos, R. and Davis, H. (1943) "The Response of
single auditory nerve fibres to acoustic stimulation" Journal
of Neurophysiology 6:39-57.

Kiang, N. Y-S. (1968) "A Survey of recent developments
in the study of auditory physiology." Annals of Otology,
Rhinology and Larynology 77:656-676.

Rose, J. E., Brugge, J. F., Anderson, D. J., and J. E.
Hind (1967) "Phase-locked response to low frequency tones in
single auditory nerve fibres of the squirrel monkey. Journal
of Neurophysiology 30:769-793.

Kiang, N. Y-S (1965) Discharge Patterns of Single Fibres
in the Cat's Auditory Nerve. Cambridge. MIT Press.

accuracy of 20 microseconds or better.

One way of plotting an aspect of the data is as the
distribution of interarrival times between successive pulses.
These are called interval histograms. One is reproduced in
Figure 5. The distribution can be described as multimodal
with the modes at integral multiples of the period of the sine
wave presented to the ear, and with the proportion of observa-
tions in each mode decaying as a geometric distribution. Thus
each active fibre has encoded the frequency of the sine wave
that is presented. Moreover, as the intensity of the sine
wave is altered, the geometric parameter changes so that the
rate increases with signal intensity. So, intensity as well
as frequency information is encoded on each fibre. Therefore,
in addition to the possibility of a place encoding of the
input we also have the possibility of a temporal encoding on
individual fibres. In the most extreme version of the tem-
poral hypothesis, one supposes that the only role of the
multiplicity of fibres -- some 30,000 of them -- is merely to
provide adequate sample sizes of interarrival times within a
relatively brief time. If, for example, we are dealing with
a 500 Hz signal, then pulses come at multiples of 2 micro-
seconds, and so on any one fibre we can expect it to take
anywhere from two to 20 microseconds to get one interval. If
the nervous system is to estimate a distribution, then
hundreds of intervals will be needed. Either this would re-
quire minutes of observation time, which is clearly not the
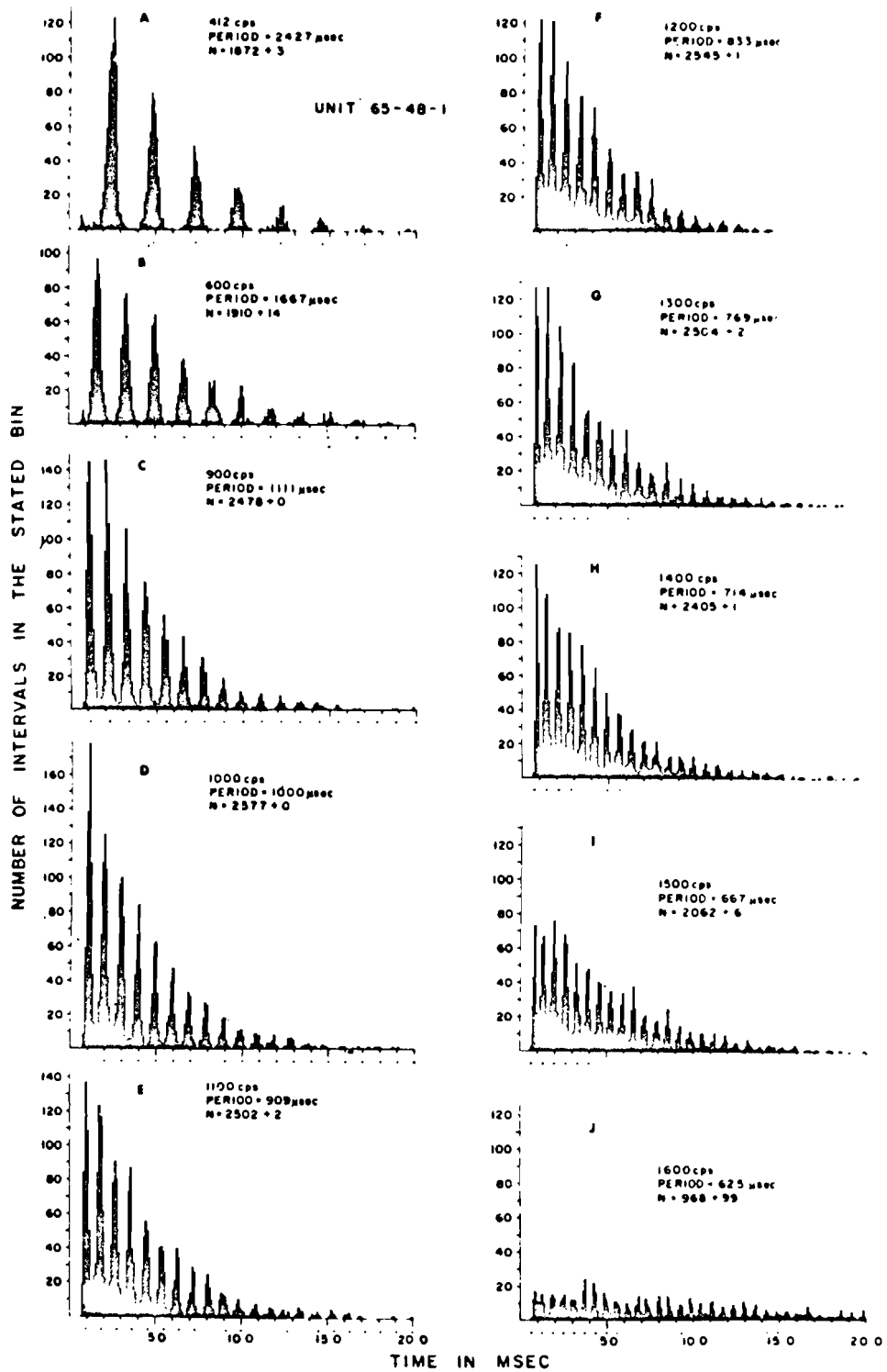case, or hundreds of parallel channels which are available.

FIGURE 5. Interarrival Times Between Successive Pulses

As yet we have said nothing about the range of fre-
quencies and intensities to which any one fibre is responsive.
Green and Luce (1973)* have described it this way:

"If we restrict our attention to pure tone signals, the
following seems to summarize the situation. Each fibre has
a characteristic signal frequency to which it is most, but
not exclusively responsive. At this frequency, there is a
lower and upper threshold. Below the lower threshold it
fires at its spontaneous rate; between the two, the rate in-
creases by a factor of 2 to 10, reaching a maximum rate at
the upper threshold;for more intense signals the rate is
either maintained or drops somewhat. As the frequency deviates
from the characteristic one, both thresholds rise and the
maximum firing rate remains about the same. Looked at another
way, a pure tone of sufficient intensity activates a particular
set of fibres in the sense that it drives their firing rates
above their spontaneous rates. Changing the frequency causes
some fibres to drop from the active category and others to
enter it; increasing the intensity adds fibres to the active
category."

One of the more perplexing questions is how the nervous
system encodes the full dynamic range of sounds -- about 12

---

*Green, D. M. and R. D. Luce (1973) "Counting and Timing
Mechanisms in Auditory Discrimination and Reaction Time," in
D. H. Krantz, R. C. Atkinson, R. D. Luce and P. Suppes (eds.)
Contemporary Developments in Mathematical Psychology, II.
San Francisco. Freedman. In Press.

orders of magnitude in intensity -- since the characteristic
frequencies of the fibres that have been studied seem to
cluster in the middle of the range. In the theory to be
described, we make a wild jump from physiological data and
hypothesize that further in from the periphery there exist
fibres where the full dynamic range is dealt with. Although
there are not yet any physiological data to support it, the
postulate is that the full dynamic range is shared by a number
of fibres, which, as a bundle, constitute a single channel
encoding the whole range. We shall assume that such a
channel exists functionally and that it, in effect, has a
dynamic range of rates between two and three orders of magni-
tude. And ignoring frequency for present purposes, we assume
that the distribution of interarrival times is a skewed geo-
metric, i.e., an exponential density with a parameter that is
an increasing function of signal intensity. Further, if we
assume that the intervals are independent, then the stochastic
process is known as a Poisson one. This is the simplest,
continuous stochastic process which corresponds to pure
temporal randomness. One way to see this is to look at the
conditional density function or hazard function: the density
of firing at time t given that no firing has occurred between
0 and t. If f is the density of interarrival times, i.e.,
$f(t) = \lambda e^{-\lambda t}$ then the hazard function h(t) is given by,

$$h(t) = \frac{f(t)}{\int_{t}^{\infty} f(x)dx} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda \; ,$$

and we see that it is uniform in time.

If this view of the encoding is approximately correct, then whenever the central nervous system (CNS) is asked to make a judgment about signal intensity it must estimate these rates in parallel channels. Two extreme ways can be suggested for doing this which are called counting and timing procedures. In the first, one fixes a time interval and counts the number of pulses that occur on each channel during that time; whereas, in the second one fixes the number of counts on each channel and times how long it takes to achieve each count. Note that in each case the CNS must measure a time, count a number of pulses, and form a ratio to get an estimate of a rate. This suggests, then, that if it can do one it should be able to do the other. So a question we can ask is whether we can demonstrate experimentally that both procedures are available. There does not seem to be any way to approach the question physiologically, so we turn to the question of demonstrating it psychologically.

Mathematical Reasoning

The problem is whether we have any chance of observing a behavioral difference. Consider the following experimental design.*   There are two tones, 0 and 1, which differ only in

---

*From this point on, the notes are taken almost verbatim from Luce, R. D. (1973) "Renewal Process Models for Psycho-physics" Social Science Working Papers, 27. School of Social Sciences, University of California, Irvine, California, 92664. It will appear as one chapter in a volume on a conference held at the Georgia Institute of Technology, June 1973, and sponsored by the Mathematical Association of America.

intensity; 1 being more intense than 0. On each trial exactly
one is presented, the schedule being random, but equally
probable. A signal is continued until the subject responds
by identifying which he thinks it is, after which there is
feedback as to the accuracy of his response and the payoff
he is to receive. The accuracy payoff o_{ij} is a sum of money
for response j to presentation i; it is positive when i = j
and negative when i $\neq$ j; varying the values o_{ij} is one way
to generate an empirical ROC curve. The deadline payoff is
simply a fine (with no payment for accuracy) whenever a
response is slower than the deadline.

At a theoretical level, let us suppose that when a signal
is presented, there are identical renewal processes on each
of the J channels. By a renewal process is meant a point
process in time such that the time between successive points
-- interarrival times (IAT) -- are independently and identically
distributed. The best known example is the Poisson process in
which the distribution of IATs is exponential; it is the model
of pure temporal uncertainty, somewhat analogous to a uniform
distribution in the finite case. Obviously, the model is
already high idealized since all of the channels are assumed
to be statistically identical, which is not true of the fibers.
Denote by M_i and V_i the mean and variance of the distribution
characterizing the renewal process for signal i, and suppose
that $M_0 > M_1$ (the more intense signal has the higher rate) and
that V_i and M_i^3/V_i are both strictly increasing functions of M_i.
(This is obviously true in the Poisson case since $V_i = M_i^2$.)

If a counting rule is used, we assume a fixed time δ (which, however, is some function of the deadline imposed and so can be manipulated experimentally) during which a count is observed. If a timing rule is used, we assume a fixed count $\kappa + 1$ per channel during which the time for κ IATs is observed. The decision must rest either on the random variable N, which is the total count over the J channels observed in time δ, or the random variable T, which is the total time for κ IATs summed over the J channels. Since the mean time between pulses is smaller for the more intense signal, it corresponds to a larger count and a smaller overall time. So plausible decision rules (which have been shown to fulfill various conditions of optimality in the theory of signal detectability) are to establish criteria, dependent upon the payoff structure, and to respond that the more intense signal was presented whenever either the count exceeds its criterion or the total time is less than its criterion, depending on which rule is in use.

Assuming that is so, let us derive the form of the ROC curve in each case. First, the counting rule. We invoke the following well known central limit theorem (Feller, 1966, p. 359)*: in a renewal process with $E(IAT) = M$ and $V(IAT) = V$, the number of counts $N(\tau)$ observed in time τ is asymptotically normally distributed with mean τ/M and variance $\tau V/M^3$, i.e.,

---

*Feller, W. (1966) <u>An Introduction to Probability Theory and its Application</u>. New York. Wiley. Vol. II.

$$\lim_{\tau \to \infty} P \left[\frac{N(\tau) - \tau/M}{(\tau V/M^3)^{1/2}} \leq z \right] = \int_{-\infty}^{z} N(0,1) ,$$

where

$$N(\mu, \sigma^2) = \frac{1}{(2\pi)^{1/2} \sigma} e^{-(x-\mu)^2/\sigma^2 2} .$$

Assuming that $J\delta = \tau$ is sufficiently large for this approximation to be good, we see that with a criterion z, we may write

$$P_{i1} \stackrel{\sim}{=} \int_{-\infty}^{z_i} N(0,1) ,$$

where

$$z_i = \frac{z + j\delta/M_i}{(j\delta V_i/M_i^3)^{1/2}}$$

is the normal deviate corresponding to the probability p_{i1}. Frequently it is convenient to represent a probability in terms of its normal deviate, especially when, as in this case, two deviates are linear functions of one another since, by eliminating c, we obtain

$$z_1 = \left[\frac{V_0}{V_1} \left(\frac{M_1}{M_0} \right)^3 \right]^{1/2} z_0 + (j\delta)^{1/2} \left(1 - \frac{M_1}{M_0} \right) \left(\frac{M_1}{V_1} \right)^{1/2}$$

as our expression for the ROC curve. By assuming V_i and M_i^3/V_i are both strictly increasing functions of M_i, it can be shown that the slope of this curve is < 1; in the Poisson case it is similar, but is based on <u>the</u> central limit theorem (Feller,

1966, p. 253) which for a renewal process with $E(IAT) = M$ and $V(IAT) = V$ asserts that the total time T required to accumulate ℓ IATs is asymptotically normally distributed with mean ℓM and variance ℓV, i.e.,

$$\lim_{\ell \to \infty} P \left[\frac{T(\ell) - \ell M}{(\ell V)^{1/2}} \leq z \right] = \int_{-\infty}^{z} N(0,1) .$$

Taking $\ell = J\kappa$, we find immediately that the ROC curve is given by

$$z_1 = \left(\frac{V_0}{V_1} \right)^{1/2} z_0 + (J\kappa)^{1/2} \frac{(M_0 - M_1)}{V_1^{1/2}} ,$$

which again is a straight line. By assumption 1, it has a slope > 1, equal to M_0/M_1 in the Poisson case. Thus, the slope of the ROC curve is a clear criterion as to which procedure is in use.

A second criterion can be found by looking at the response times. These times are the sum of two parts, that taken up accumulating information about the signal, which we call the decision time, and all other times, including those for computations, transmissions, muscle movements, etc., which we call the residual time. Let the mean of the residual time be denoted by $\bar{r}$, then for the counting model the mean response time for signal i and response j is

$$MRT_{ij} = \bar{r} + \delta .$$

The prediction is that it is independent of i and j; indeed, if the decision and residual times are independent, it is not just the mean, but the whole distribution that is predicted to be independent of i and j. In the timing model, the decision time is more complicated because it is determined by the slowest of the J channels to observe κ IATs. Denote by $h(J,\kappa,\sigma)$ the mean of this time when the renewal process has mean 1 and variance σ^2. Then, for a process with mean M_i and variance V_i,

$$MRT_{ij} = \bar{r} + h(J,\kappa,V_i^{1/2}/M_i)M_i$$

In the Poisson case, $V_i^{1/2}/M_i = 1$, and so for it and any other case in which this ratio is nearly constant we may eliminate h and write

$$MRT_0 = \left(\frac{M_0}{M_1}\right) MRT_1 - \bar{r}\left(\frac{M_0}{M_1} - 1\right) ,$$

where we have dropped the response subscript j since, by the next to last equation, it does not matter. So, as we vary the deadline, we should find a linear relation between the mean response times; moreover, the slope should be identical to that of the corresponding ROC curve.

Green and Luce (1973a) ran three observers in such an experiment using a faint 1000 Hz tone in noise for 1 and noise alone (0 intensity signal) for 0. The design was as described, with deadlines varying from 250 msec to 2000 msec. When the deadline applied to all trials, the mean response time was the

same in all four cells except for the two longest deadlines,
where there was some tendency for the signal trials to be
slower than the noise ones; we return to this discrepancy
below. The ROC data (Fig. 6) were well fit by straight lines
with estimated slopes of 0.92, 0.69 and 0.90, all supporting

Figure 6

ROC Curves for three observers from a Yes-No experiment
involving a signal in noise (s) versus alone (n), i.e.,
Y = s = 1, No = n = 0 .

The coordinates are normal deviates. (This is Fig 4 of
Green and Luce 1973a).*
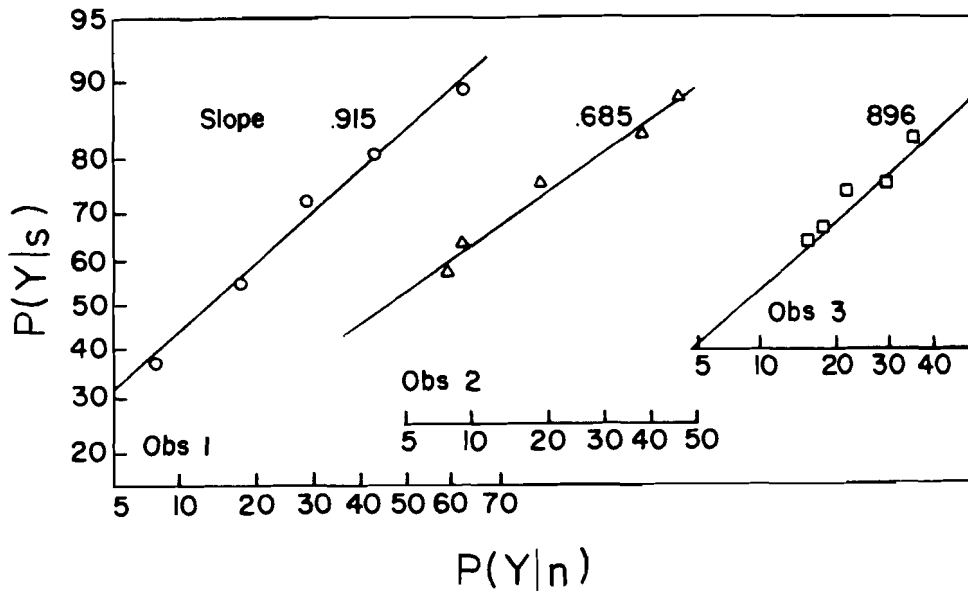
*Green, D. M. and Luce, R. E. (1973a)"Speed-accuracy
trade-off in auditory detection" in S. Kornblum (ed.)
Attention and Performance IV. New York. Academic Press.

the counting model. When three other observers were run in exactly the same experiment except that the deadline applied only to signal trials, both the MRT and ROC data were well approximated by straight lines and the pairs of estimated slopes were:

1.34, 1.30; 1.48, 1.47; and 1.38, 1.37.

The timing model was clearly supported.

A striking way to show up the differences between the models and between these two sets of data is as a trading relation between speed and accuracy. A very common measure of accuracy, suggested by the theory of signal detectability, is to compute the value of z_1, called d', corresponding to $z_0 = 0$. For the counting model,

$$d' = A\delta^{1/2} \, ,$$

and for the timing model

$$d' = \frac{M_0}{M_1^{1/2}} \, A\kappa^{1/2} \, ,$$

where

$$A = J^{1/2} \left(1 - \frac{M_1}{M_0}\right) \left(\frac{M_1}{V_1}\right)^{1/2} \, .$$

Eliminating δ yields

$$d' = \begin{cases} A(MRT - \bar{r})^{1/2} & MRT \geq \bar{r} \\ \\ 0 & MRT < \bar{r} \end{cases}$$

for the speed-accuracy trade in the counting model.

For the timing model, the equation for mean reaction time must be developed more fully before we can eliminate κ. If we let ψ_κ denote the distribution of the κ IATs, then by definition

$$h(J,\kappa,\sigma) = J \int_0^\infty x\psi_\kappa(x) \left[\int_0^x \psi_\kappa(y) \, dy \right]^{J-1} dx .$$

Assuming that ψ_κ is approximately normal, which for $\kappa \geq 5$ is not a bad approximation in the Poisson case, then it follows readily that

$$h(J,\kappa,\sigma) \simeq \kappa + 1 + (\kappa+1)^{1/2} H(J) ,$$

where $H(J)$ is the mean of the largest of J random variables distributed $N(0,1)$. If we substitute this into the equation for mean reaction time and introduce the variable

$$T_i = MRT_i - \bar{r} - M_i - V_i^{1/2} H(J),$$

and the constant

$$B_i = V_2^{1/2} H(J)/M_i ,$$

then eliminating κ yields the speed-accuracy trade

$$(d')^2 = \begin{cases} \dfrac{M_0^2}{M_1 M_i} \ A\left\{ T_i + \dfrac{M_i B_i}{2} \left((B_i+2) - [(B_i+2)^2 + \dfrac{4T_i}{M_i}]^{1/2} \right) \right\} & T_i \geq 0 \\ \\ 0 & T_i < 0 \end{cases}$$

Three qualitative differences can be seen by comparing the two equations that give d'. First, the last point for which d' = 0 is larger in the timing model than in the counting

one by the amount $M_i + V_i^{1/2} H(J)$. Second, because the times

MRT_1 and MRT_2 are different in the timing model, there are

two trading relations. Third, the initial slopes of the

speed-accuracy trade equation for the timing model are greater

than that of coefficients for d' in the counting case by

factors $(M_0/M_1)^{1/2}$ and M_0/M_1. Figure 7 shows the data plotted

in this way, with the data for the observers combined in the

Figure 7.

Plots of d' versus MRT for Two Experimental Conditions
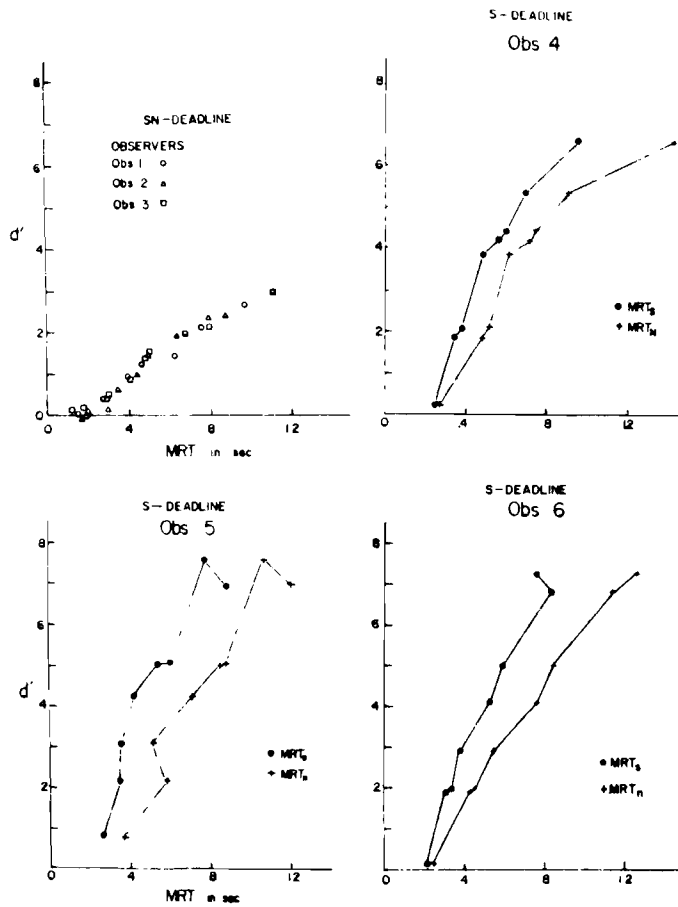
first experiment and separated for clarity in the second
one. We see that the qualitative predictions are sustained.
In fact, the growth of the function in the second experiment
is so much more rapid than in the first that the accuracy at
long deadlines is considerably greater using the timing rule
than the counting one. The evidence suggests that two of
the three observers realized this and switched to timing be-
havior for the long deadlines, thus producing the discrepancy
in times mentioned earlier.

Magnitude Estimation

Another set of related results has been suggested by
some data collected by the lecturer and D. M. Green using
the method of magnitude estimation. In this method, due to
S. S. Stevens, subjects are asked to associate positive
numbers to (acoustic) stimuli in such a way that the ratios
of the numbers represent the subjective ratios of the sounds
presented. Although the technique might sound implausible,
in fact, highly regular and repeatable results are obtained.

If we assume that subjects have an internal representa-
tion of signal intensity, such as proposed by either the
timing or counting models, then a possible interpretation
of our instructions is for the subjects to make the ratio
of the present response to the previous one equal to the
ratio of the representations. This is the most obvious in-
terpretation the subjects could place on the instructions
and it suggests that there should be strong sequential

dependencies upon the preceeding signal response, as there are (Ward, 1973).[*]

In our theoretical calculation, let us suppose that the rate estimated according to the timing decision rules applied to a Poisson process is the subject sense of the signal. Let μ be the intensity parameter of the Poisson process, k (=Jκ) be the IAT sample size, T the random variable consisting of the sum of all IATs observed, and R the random variable equal to the subject's numerical response. A prime added to any symbol simply refers to the trial preceeding the one for which an unprimed symbol is used. Our response hypothesis, then, is simply

$$\frac{R}{R'} = \frac{T'/k'}{T/k} .$$

By elementary distribution arguments, we see that

$$\text{where} \quad P(\frac{R}{R'} = x) = P(\frac{T'k}{T k'} = x)$$

$$P(\frac{T'}{T} = y) = \frac{1}{B(k,k')} \frac{y^{k'-1}(\mu'/\mu)^{k'}}{(1 + \mu'y/\mu)^{k+k'}} , \tag{1}$$

and

$$B(k,k') = \frac{(k-1)!(k'-1)!}{(k+k'-1)!} .$$

Equation (1) is the beta distribution of the second kind. Its mean m and variance σ^2 are readily calculated, from which

[*]Ward, L. M. (1973) "Repeated Magnitude Estimates with a Variable Standard: Sequential Effects and Other Properties." Perception and Psychophysics 13:193-200.

$$\left(\frac{\sigma}{m}\right)^2 = \frac{k + k'-1}{k'(k'-2)} \tag{2}$$

Observe that the right side is independent of the rate parameters μ and μ', but since $k = J\kappa$ may depend on intensity (through J) the ratio can still be a function of intensity.

In data from several observers who responded approximately 500 times to each of 20 signals spaced equally in dB over a 50 dB range, for any given intensity ratio (equal dB difference) there is no evidence that σ/m changes greatly as the absolute level of intensity changes. This suggests that J is nearly independent of intensity and so $k \simeq k'$. For each intensity ratio we, therefore, average all of the data over the different levels of intensity, obtaining for one observer the plots of m and σ/m versus intensity shown in Figure 8. The first thing to note is that m grows approximately as a power function of intensity (see next section for further evidence on this point). Second, although σ/m is probably a constant for signal ratios in excess of 20 dB, it decreases by a factor of as much as 3 for smaller ratios. Assuming $k = k'$ in each case, we get estimates of sample sizes of roughly 19 and 167 respectively. The former seems very small.

The question is whether we can make theoretical sense of these data. One idea, which although it has not yet been worked out in detail seems to have the correct qualitative features, is that the CNS is able to focus only on a limited

Figure 8. Mean and Coefficient of Variation (σ/m) of the
ratio of successive magnitude estimates as a function of
the ratio of the stimulus intensities for one observer.
(Unpublished data of D. M. Green and R. D. Luce.)
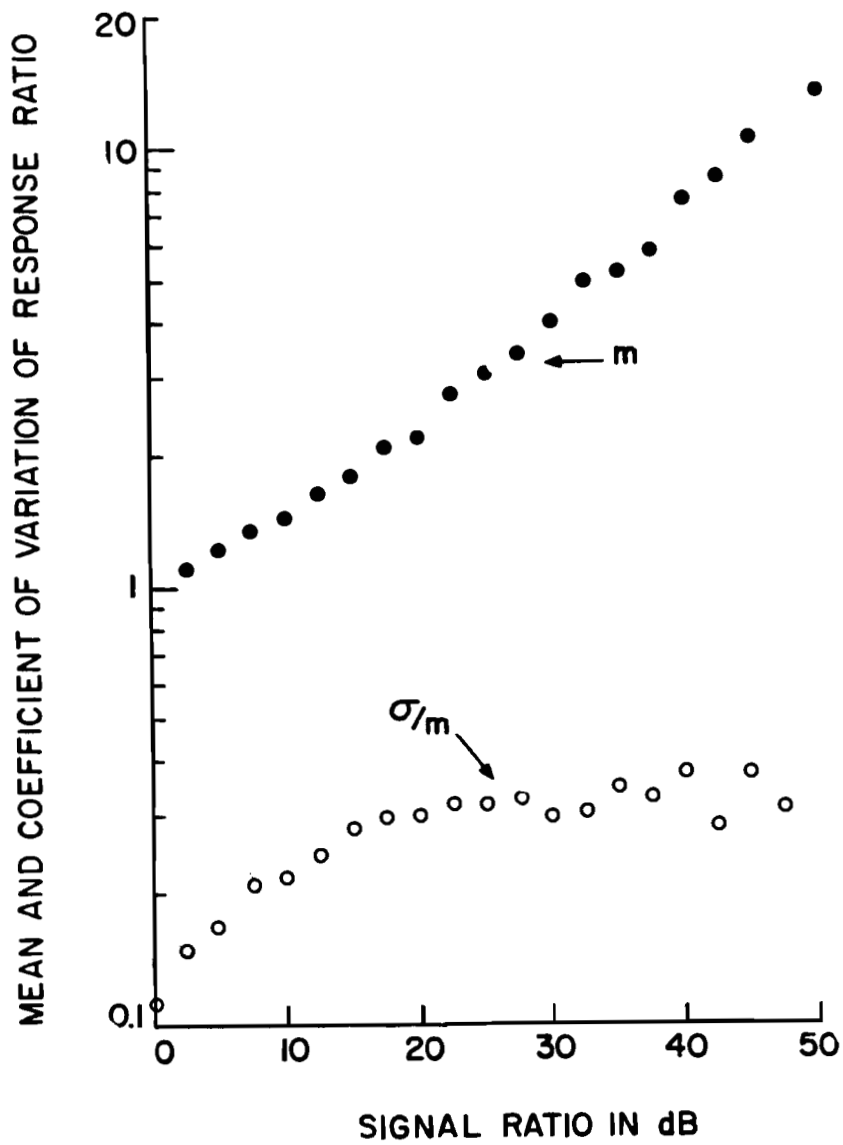
range of intensities at any one time. An extreme version of
this model supposes that it can collect a large sample, say
167, only for rates falling in a range corresponding to about
20 dB, and for rates outside that range, only a much smaller
sample, say 19, is possible. This makes sense if we postulate
that each nerve fiber has only dynamic range of about 20 dB,
as seems to be the case, and that as intensity is changed
some fibers are saturated while others are brought into play.
The assumption would then be that the CNS can monitor fully,
with its largest sample size, only these fibers corresponding
to one 20 dB range, and that activity outside that range is
monitored only with much smaller samples. This assumption
makes the 20dB limit in the physiological data account for
the 20 dB edges in this magnitude estimation experiment and
in certain absolute identification ones. Furthermore, it
suggests experiments to test itself. If we can manipulate
the range that the CNS is monitoring, then we should get
predictable phenomena. For example, by introducing sequential
dependencies in the signal presentation schedule, we can make
the probability that two successive signals are within 20 dB
of each other as low or high as we choose, which should affect
the tendency of the subject to monitor near or far from the
value of the previous signal. If we then compare the behavior
to the exceptional signals -- the near ones when near ones
are improbable and the far ones when far ones are improbable
-- with the behavior to the common ones, we should find the
former much more variable than the latter.

Response Time to the Onset of a Signal

Our last application illustrates the fact that problems that are simple to formulate in this Poisson theory do not necessarily lead to solved mathematical questions.

Consider an experiment in which a signal comes on at a random time after a warning signal, and the observer is to respond to it as rapidly as possible without, however, making too many anticipatory responses. Formally, the data from each trial consist of a pair of random variables S and R, where S is the time at which the signal (say, a change in intensity) comes on and R is the time at which the sugject responds. Denote by

$$f(x,t) = P(R = t | S = x)$$

the conditional probability density that the response time is t when the signal onset time is x, and by

$$g(x) = P(S = x)$$

the density of the signal onset times, which is under experimental control.

In terms of the model, we suppose a Poisson process with parameter ν until S and one with parameter $\mu (> \nu)$ after S. Some sort of decision rule will be applied to this stochastic process, leading to a decision at time D (< R) to initiate the response. Whatever that rule may be, let

$$\ell(x,y) = P(D = y | S = x)$$

denote the conditional probability density that the decision time is y when the signal onset time is x. We refer to the

time R-D as the <u>residual</u> <u>time</u> -- it consists of all the times
consumed by the nervous system aside from those taken up in
arriving at a decision. We make the following assumptions
about this random variable:

  (i) R-D and D are independent random variables.

  (ii) R-D and S are independent random variables.

  (iii) R-D is a bounded random variable.

Empirically, there is some doubt whether (i) and (ii) are
strictly correct. For example, the readiness to respond may
be affected by the overall delay, and so by the value of S.
The evidence in favor of (iii) is the boundedness of response
times to intense signals. The bound appears to be the order
of 300 msec. By (ii) one can reasonably postulate a density
for R-D, call it r, and by (i) we see that

$$f(x,t) = \int_0^t \ell(x,y)r(t-y)dy.$$

It is convenient to divide the observable response time into
two parts corresponding to anticipatory responses and those
that appear to be in response to the signal, specifically

$$f_R(t) = \frac{\int_t^\infty g(x)f(x,t)dx}{\int_t^\infty g(x)dx}$$

$$f_{R-S}(t) = \frac{\displaystyle\int_0^t g(x)f(x,t)dx}{\displaystyle\int_0^t g(x)dx} \quad .$$

In the experimental analysis and data given below, the onset density was exponential,

$$g(x) = \lambda e^{-\lambda x} \quad .$$

The reason for this choice is that it makes ineffective any possible strategic considerations in responding based on how long the subject has waited.

To the best of my knowledge, the following basic question has not been answered: given a payoff function $P(S,R)$ (where, presumable, $P(S,R) < 0$ for $R < S$), what is an optimal decision rule to detect a simple increase (or decrease) in the parameter of a Poisson process? The answer to this might provide some suggestions about the sort of rules employed in the CNS.

A far simpler question, although not without difficulties, is to postulate the simplest rule one can think of, derive properties of ℓ, f_R and f_{R-S}, and compare the latter two with data. The simplest rule we have thought of (see Luce & Green, 1972)*, which is both the most responsive and most variable

---

*Luce, R. D. and D. M. Green (1972) "A Neural Timing Theory for Response Times and the Psychophysics of Intensity." Psychological Review 79:14-57.

way to detect a change, is to select a criterion β and com-
pare each IAT with it, initiating a response the first time
after the warning signal that IAT < β. The value selected
for β will depend, of course, on the magnitude of the change
to be detected and the payoffs. Assuming this rule and
denoting by ℓ_0 the density ℓ conditional on a pulse at time
0, elementary probability considerations lead to the follow-
ing pair of integral-difference equations for ℓ and ℓ_0

$$
\ell(x,t) = \begin{cases} \displaystyle\int_0^t \nu e^{-\nu y}\ell_0(x-y,t-y)dy, & t < x \\[4mm] \displaystyle\int_0^x \nu e^{-\nu y}\ell_0(x-y,t-y)dy + e^{-\nu x}\int_x^t \mu e^{-\mu(y-x)}\ell_0(0,t-y)dy, & t \geq x \end{cases}
$$

$$
\ell_0(x,t) = \begin{cases} \nu e^{-\nu t}, & t < x \\[2mm] e^{-\nu x}\mu e^{-\mu(t-x)} & x \leq t \leq \beta \\[2mm] e^{-\nu x}e^{-\mu(\beta-x)}\ell(0,t-\beta), & x < \beta \leq \\[2mm] \nu e^{-\nu t} & t \leq \beta \\[2mm] e^{-\nu\beta}\ell(x-\beta,t-\beta), & \beta < t \end{cases}
\left. \begin{array}{} \\ \end{array} \right\} x < \beta \quad\quad \left. \begin{array}{} \\ \end{array} \right\} x > \beta
\tag{3}
$$

The technical problem is to solve these equations. Al-
though this has not be done fully, enough is known to suggest
that the model is not wholly absurd. In particular, for
sufficiently large t, the boundedness of the residual times
plus the fact that a slow exponential decay dominates the
solution permits one to show that

$$f_R(t) \simeq Ae^{-(\lambda+\nu')t} \; ,$$

where

$$\nu' = \nu(1-e^{-(\nu-\nu')\beta}) \; ,$$

and

$$f_{R-S}(t) \simeq Be^{-\mu't} \; ,$$

where

$$\mu' = \frac{\mu^2\beta}{1 + \mu\beta} \tag{4}$$

and A and B are some functions of the parameters. Of course, the tails of the distributions constitute only a fraction of the data and we would really like to know the form of the entire solution, but at present we have no option but to waste much of the data.

A first test of the model is to see if the tails of the distributions are approximately exponential. A sample of data is shown in Figure 9, for times greater than 1/2 sec, and the approximation is not bad. Using the previous four equations to estimate ν' and μ' for different intensities yields Figure 10. Observe that the growth of μ'/ν' is approximately a power function of intensity, which agrees with the conclusion from ME data. However, there is a considerable discrepancy in the estimates of the exponents: something of the order of 0.30 ± 0.15 from the ME data for a variety of subjects and about 1.5 from these reaction time data.

As Luce and Green (1972) pointed out, one source for

Figure 9

Tails (t ≤ 0.50 sec.) of the false alarm and reaction
time distributions for one observer at one intensity
level. The smooth curves are the best fitting exponen-
tials, and the inserts present the data when grouped in-
to 20 equiprobable intervals according to those distri-
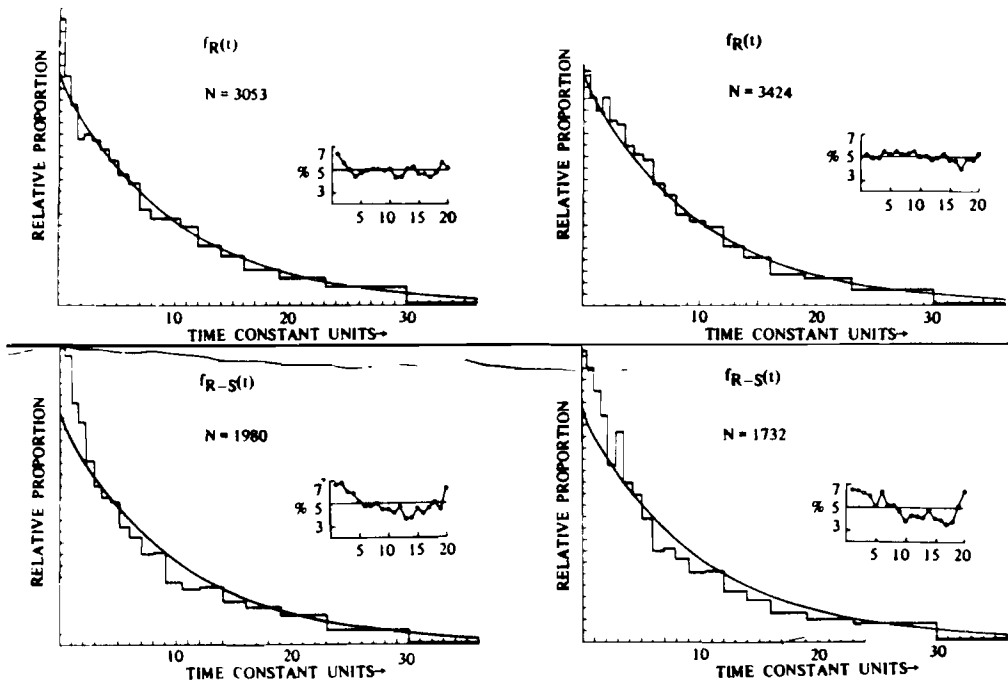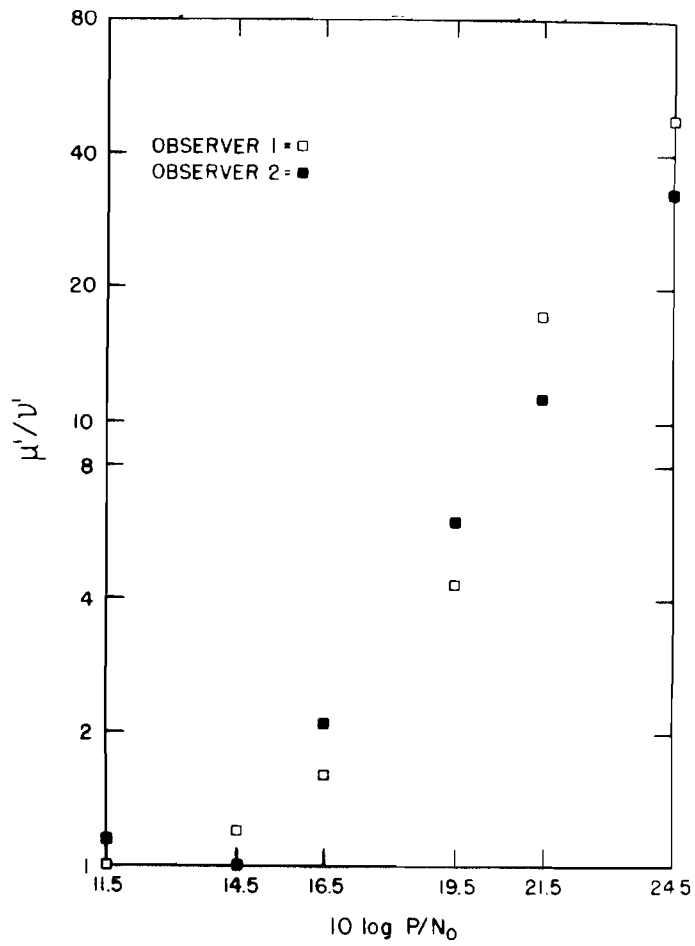butions.

Figure 10

Ratio of Poisson parameters estimated from the tails of
the reaction time distributions from two observers as
a function of signal-to-noise ratio in dB   (This is Fig.
9 of Luce and Green, 1970).

the discrepancy is that the reaction time analysis is for
only a single channel, and the following model suggests that
this may be the only source. Suppose that the single channel
analysis applies independently to each of J parallel channels,
each with a criterion δ, and that a decision does not initiate
a response, but rather causes another channel to fire. This
common channel, which receives inputs from each of the J
channels, applies the same decision rule, but with criterion
β, and it initiates a response whenever two of the J channels
fire sufficiently closely.

If the rate of the underlying process is μ, then the
mean IAT of the decision process is

$$\frac{1}{\mu(1-e^{-\mu\delta})} ,$$

as can be shown by deducing the Laplace transform of $\ell_0(0,t)$
from Equation (3) and then calculating the mean in the usual
way. By a well known theorem (Cox, 1962, p. 77-79)* the
superposition of J independent, identical renewal process
approaches a Poisson process as $J \to \infty$; moreover, its rate
is given by

$$\mu^* \simeq J\mu(1-e^{-\mu\delta}).$$

If we assume that both δ and β are sufficiently small so
that we may use linear approximations to the exponentials in
the preceding equations and drop the term $\mu\beta$ in Equation (4),

---

*Cox, D. R. (1962) Renewal_Theory. London, Methuen.

$$\frac{\mu'}{\nu'} \simeq \frac{(\mu^*)^2 \beta}{1+\mu^* \beta} \cdot \frac{1+\nu^* \beta}{(\nu^*)^2 \beta} \simeq \left(\frac{\mu^*}{\nu^*}\right)^2 \simeq \left(\frac{\mu}{\nu}\right)^4 \ ,$$

which neatly accounts for the factor of 4 discrepancy between the ME and reaction time estimates.

Concluding Remarks

I hope that the following points have become clear as a result of my illustrative psychophysical models.

1. Psychology has empirical problems of some complexity which can be significantly illuminated by using probability models accessible to undergraduate students.

2. As in physics, radical oversimplifications (identical channels, the simplest of decision rules, Poisson processes) of the micro-structure (neural pulse trains) can, if handled with care, provide adequate qualitative and even quantitative models of the macro-structure (psychophysics).

3. Courses on stochastic processes for social science majors probably should include some material on continuous-time stochastic processes, especially Poisson processes and perhaps more general renewal ones. Psychologists are generally less familiar with such processes than with dis-crete time ones (especially Markov chains), and as a result they have developed little theory for situations in which responses can occur at any time (e.g., Skinnerian operant conditioning experiments) and they tend to employ experimental designs with a trial structure, which may very well seriously

distort an organism's performance from what it would be under more natural conditions. Moreover, as I have tried to demonstrate, continuous time processes are probably satisfactory models for some neural activity and certainly can serve as idealized underpinnings for psychological theories.