

# WHAT SORT OF MEASUREMENT IS PSYCHOPHYSICAL MEASUREMENT?<sup>1</sup>

R. DUNCAN LUCE<sup>2</sup>

*The Institute for Advanced Study, Princeton, New Jersey*

THE thesis of this article has both a negative and a positive aspect. The negative one holds that psychophysical measurement is not of a character closely analogous to either fundamental or derived physical measurement. Whatever loudness may be, I endeavor to establish that it probably is not a measure much like mass or energy or density. In brief, the reason is that psychophysical measures do not exhibit any fixed relation to physical measures and most likely not to one another when examined over individuals. This is reflected in the absence of any structure to the units of psychophysical measures.

The positive thesis is that man—and any other organism—is, among other things, a measuring device, in function not unlike a spring balance or a voltmeter, which is capable of transforming many kinds of physical attributes into a common measure in the central nervous system. According to this view, the task of psychophysics is to unravel the nature of that device. The difficulty in doing so stems from several facts: as compared with man-made, special-purpose devices, higher organisms are both complex and flexible measuring devices; their overall behavior does not clearly suggest the nature of the recoding of signals; some responses depend on peculiar nonlinear processing of the sensory information; and each individual within a species is calibrated somewhat differently.

To flesh out and make this view reasonably clear, we must know something of physical measurement and of psychophysics. Since the details of the nature of physical measurement, especially the abstract theory, are probably not very familiar

to most psychologists, I devote some space in this article to them.<sup>3</sup> To keep these remarks within bounds, I am forced to assume a general familiarity with contemporary psychophysics, to the extent that mere mention of a method, theory, or result is sufficient to bring it to mind. The remainder of the article is devoted to showing how poorly the fundamental measurement concepts of physics fit the structure of psychophysics and how much more satisfactory is its interpretation as the study of a measuring device.

## *Brief Comments on the Nature of Physical Measurement*

An essential distinction in understanding physical measurement is that between theory and practice—even though practice often rests on a good deal of sophisticated physical theory beyond measurement theory. Put another way, this is the distinction between the foundations of measurement, that is, the study of which ordinal attributes can be represented systematically by which numerical structures, and the development and calibration of devices designed to render such measurement convenient and, in some cases, practicable. A simple example will make the distinction transparent. The fundamental analysis of weight measurement is based on having a very large, very refined collection of objects and an idealized, equal-arm pan balance. Placing a finite collection of objects in each pan and noticing which one drops permits us, in principle, qualitatively to order by weight all finite collections of objects. Certain empirical generalizations are so familiar as to seem tautological, even though they are not. The most important is that if A, B, C, and D are mutually disjoint collections such that A is at least as heavy as C, and B is at least as heavy as D, then the union of A and B will be found to be at least as heavy as the

<sup>1</sup> The preparation of this article was supported by a grant from the Alfred P. Sloan Foundation to The Institute for Advanced Study. It was delivered as a Distinguished Scientific Award Address at the annual meeting of the American Psychological Association, September 1971, Washington, D.C.

<sup>2</sup> Requests for reprints should be sent to R. Duncan Luce, The Institute for Advanced Study, Princeton, New Jersey 08540.

<sup>3</sup> Anyone interested in a fuller discussion of these matters should consult Krantz, Luce, Suppes, and Tversky (1971), especially Chapter 10.

union of C and D. With a suitable set of such generalizations as axioms, a classical mathematical theorem (Helmholtz, 1887; Hölder, 1901; for modern formulations see Chapters 2 and 3 of Krantz et al., 1971) shows how to assign numbers to the objects in such a way that sums over finite collections of them order them numerically in exactly the same way as does the pan balance.

This, then, is the basic theory of weight measurement, and any proposal for the actual weighing of objects must be shown, in principle, to be equivalent to it. In practice, pan balances are not terribly convenient, although they were common enough until well into this century and can still be found in rural areas all over the world, and so other devices such as spring balances are often used. Such a balance is based on the empirical fact that forces systematically alter the length of a spring. So the measurement of weight is transformed into an equivalent one of length, which is rather easier to measure. The only remaining step is to calibrate the device, that is, to determine from fundamental measurement of weight and of length exactly how length covaries with weight in the particular spring balance.

In general, then, a measuring device is any specially constructed (or naturally occurring) apparatus that transforms one attribute into another in an (approximately) one-to-one fashion. It is a useful device if we wish to measure one of the attributes and the other one is easy to measure. Often length is the one that is easy to measure. It is, of course, essential to know how to measure each attribute independently of the device in order to be able to calibrate it. This last point is sometimes overlooked by those who introduce so-called operational definitions of attributes. The lack of a deep measurement analysis of anxiety is not overcome by saying that, in a particular context, anxiety is by definition skin resistance measured in ohms at a particular location on the body. A clue that something is missing is the failure to discuss the calibration of the ohm meter in nonarbitrary units of anxiety.

Next, let us consider something of the structure of the theory of physical measurement. One aspect is that it provides a way to pass from an ordinal definition of an attribute to a numerical representation of it. There are now a number of results along these lines, the two most important for

physics being extensive measurement, such as describes the measurement of weight, and additive conjoint measurement (Debreu, 1960; Luce & Tukey, 1964), which, for an attribute having two or more independent components that affect it, formulates conditions that are sufficient to construct a numerical representation that is additive over those components.<sup>4</sup> Examples of the latter are the dependence of kinetic energy on mass and velocity and that of mass on volume and substance.

Much emphasis has been placed on the numerical representation of isolated physical attributes, often, I fear, at the expense of overlooking a second equally important feature of physical measurement. The several—well over 100—attributes of physics are locked together, forming a rigid algebraic structure which embodies a number of simple, basic physical laws. It is the existence of this structure which makes the method of dimensional analysis work. Its most obvious manifestation is the familiar, if ill understood, composition of units. At present, the structure is six-dimensional in the sense that six attributes can be chosen independently—one suitable set is length, mass, time duration, angle, temperature, and charge—and all of the others are monomial functions of these. This shows up in the fact that the units of all other physical measures can be taken to be of the form

$$\text{meter}^\lambda \text{ gram}^\mu \text{ second}^\sigma \text{ radian}^\epsilon \text{ degree}^\theta \text{ coulomb}^\chi,$$

where the exponents  $\lambda$ ,  $\mu$ , etc., are rational numbers. This is truly quite remarkable and bears some analysis.

The key to constructing this algebra is the existence of attributes that are both extensively and conjointly measurable. The conjoint theory, when it applies, constructs interlocked measures on three attributes. It is a fact of physics that at least two of the three attributes involved in any conjoint structure are also extensively measurable; that is, empirical operations of addition are possible on two of them. This means, then, that at least two-thirds of these attributes have two quite distinct measures defined. The one is based on an internal, additive structure of the attribute and the other on its relation to two other attributes. Of course, these duplicate measures are monotonically related to each other since they preserve the same

<sup>4</sup>In physics, it is customary to write these conjoint measures in multiplicative rather than additive form—the only difference being an exponential transformation.

ordinal structure, but beyond that the measurement theories alone cannot take us. If, however, certain types of qualitative empirical laws hold—they are called laws of exchange and of similitude by Krantz et al. (1971)—then these monotonic functions are actually power functions. This means that the measure of the nonextensive attribute of a conjoint triple can be written as a product of powers of the two extensive measures. Assuming that sufficiently many such laws hold, as they do in physics, one can show that the resulting algebra has the form that the measures of all attributes can be expressed as products of powers of the measures of a finite, basic set of extensive attributes.

Those nonextensive attributes that enter into a law of similitude or exchange can be expressed in terms of extensive ones, and so they are often called derived. An example is density. In this case, mass is the conjoint attribute which can be manipulated by varying volume and substance independently. Both mass and volume are extensive attributes. The law of similitude that holds is a qualitative analogue of the usual numerical statement that, for a given substance, the ratio of mass to volume is independent of the volume. That ratio, which is the conjoint measure associated to substance, is called density. From one point of view it is misleading to call such attributes derived because they are measured fundamentally via conjoint measurement procedures; however, they are not extensively measured, and they can be “derived” from other extensive measures via a law of exchange or of similitude.

#### *A Brief Summary of Psychophysical Measurement*

Three quite distinct bodies of theory and data from psychophysics have been interpreted as forms of measurement.

1. There are attempts to construct numerical representations of ordinal data; often this is called scaling. Perhaps the best known example is Shepard's (1966) very successful program for representing ordinal judgments of the similarity of stimuli as distances in a Euclidean space of minimum dimension. A second example is the current attempt to scale such attributes as loudness by means of additive conjoint measurement (Levelt, Riemersma, & Blunt, in press). A third is Pfanzagl's (1959, 1968) proposed theory of bisection which, to my knowledge, has never been applied to data.

Still other examples can be found in Coombs' (1964) book.

2. There are attempts to treat numerical responses to signals as numerical measures of attributes. Perhaps the most widely used method of this type in psychology is the rating scale, but far more important for psychophysics is Stevens' (1957) method of magnitude estimation and its close relatives, magnitude production and cross-modality matching. The striking regularities of the data obtained by magnitude methods—not only the familiar power functions (Stevens, 1957; Stevens & Galanter, 1957) and the predicted average exponents of the matching work (Stevens, 1959, 1966b), but also the representation of intensity-duration exchanges (Stevens, 1966a; Stevens & Hall, 1966), the effect of adaptation (Stevens & Stevens, 1963), and the close relation to reaction time (McGill, 1963; Vaughan, Costa, & Gilden, 1966)—have impressed many of us and should challenge theoreticians to provide theories that encompass both these methods and the more traditional ones having to do with discrimination and detection.

3. There are attempts to interpret estimated parameters of probabilistic models as measures of psychological attributes. This, the classical and most widespread, approach to psychophysical theorizing can be traced back at least to Fechner's (1860) famed rescaling of just noticeable differences, which in recent times has been reappraised by Luce and Edwards (1958) and Falmagne (1971), generalized in Levine's (1970, in press) studies of uniform and similar systems of functions, and modified by Eisler (1963). A second type of theory in which parameters can be interpreted as measures includes Thurstone's (1947) law of comparative judgment for paired-comparisons data, its generalizations to other kinds of data (Bock & Jones, 1968; Coombs, 1964), and its close relatives, such as the BTL model (Bradley & Terry, 1952; Luce, 1959). Related models for detection experiments include the theory of signal detectability (Green & Swets, 1966; Tanner & Swets, 1954) and the low threshold model (Luce, 1963). Perhaps the best known example of such a parameter is  $d'$  from detection theory which, because it increases systematically with signal-to-noise ratio and because it is unaffected by payoffs and various other changes in experimental design, is interpreted as a measure of the detectability of the signal. Still

another type of theory postulates a recoding of signal intensity into neural pulse trains. The pulse rate is an inferred parameter which is interpreted as a psychological measure of intensity. Considerably more, including references, is said below about such models.

To summarize, psychophysical measurement can be grouped into the scaling of ordinal data, magnitude methods, and the estimation of parameters from probabilistic models.

### *Does Psychophysical Measurement Resemble Physical Measurement?*

Current wisdom about psychological measurement, most forcefully and carefully articulated by Suppes and Zinnes (1963), sees the following analogy to physical measurement.<sup>5</sup>

1. The attempts to pass, by means of a formal axiomatic theory such as extensive or conjoint measurement or certain other measurement theories not used in physics, from ordinal data to numerical representations of these data are attempts at fundamental measurement, comparable to those successfully carried out in physics. Whether any of these attempts in psychology is yet successful is debatable, but according to this view, if some axiomatic measurement theory is shown to fit a psychological attribute, then that attribute has been measured in the same fundamental sense as in physics. The work of Campbell and Masterson (1969) on aversiveness to electric shock in rats and of Levelt et al. (in press) on the loudness of pure tones in human beings encourages one to think that some psychological attributes may fulfill the conditions of additive conjoint measurement.

2. Magnitude estimation can be seen simply as a special case of cross-modality matching in which the match is between the number scale and the physical stimulus. Furthermore, as Krantz (in press) demonstrated, cross-modality matching may possibly satisfy ordinal axioms that lead to a numerical representation in which each attribute is

<sup>5</sup> Savage (1970), in an extended critique of psychophysics, argued that psychophysical measurement should not be viewed in the following way. Roughly, he attempted to show that loudness and pitch are really physical measures and that the business of psychophysics is the study of perceptual abilities. For a brief, critical review of his position, see Luce (1971).

some fixed power function of each of the others. If so, magnitude estimation is then no longer a special category of psychophysical measurement because it has been subsumed as a part of fundamental ordinal measurement.

3. The probabilistic models from which parameters are estimated are, in reality, just empirical laws, comparable to those of exchange and similitude in physics, and so the parameters are simply derived measures, comparable to density. Of course, the parameters are expressed in terms of response probabilities, but since probability is a kind of extensive measurement, the analogy to physics is said to be quite close—the derived measures are reduced to prior extensive measurement.

All of this is true, in a sense, and yet I claim that psychophysical measurement is really not very much like physical measurement. A key difference is seen when we ask what is the algebraic structure relating the psychophysical measures; it should be reflected in the pattern of their units. It is conceivable that the set of psychophysical attributes has a structure itself. If so, it surely will not be constructed in the same way as the physical structure, that is, using extensive and conjoint measures interlocked by qualitative laws, because no psychophysical attribute has been shown to be extensively measurable. But other structures are possible, for example, the one resulting from cross-modality procedures. For this to work out so that all of the units can be expressed as monomial functions of a few, it will be necessary for the exponents relating attributes to be independent of the subject. Although I know of no data to exclude this possibility, I find it difficult to be optimistic that it will hold. Clearly, however, Krantz's theory must be checked empirically, and, in particular, it must be decided whether or not the exponents are invariant over people.

An alternative approach, for which we do have the data, is to suppose that the psychophysical measures are to be adjoined to the structure of physical ones. Were this to work out, loudness, for example, would be expressed as products of powers of certain physical measures—among them, sound intensity—and its unit would be a product of (rational) powers of the six basic physical units and perhaps some psychological units. As a matter of fact, there is considerable evidence from magnitude estimation that loudness is (approx-

mately) a power function of intensity, and Levelt et al. (in press) have shown the same result using conjoint procedures on the two ears. The only trouble is that the exponents vary from subject to subject—over a range from at least .15 to .50—and even between the two ears of a single subject. Thus, no unit of loudness can be stated, independent of the individual, in terms of physical units.

In sum, it seems to me that it is incumbent upon those who allege a similarity of psychophysical measurement to fundamental physical measurement to show the existence of invariant relations among the measures and to develop a coherent system of units as has been done in physics. This has not been done, and I doubt that it can be.

Let us next turn to the resemblance of inferred parameters of models to derived measures of physics. I claim that three differences are sufficiently marked to make that resemblance remote. First, the type of law formulated by a theory such as signal detectability theory is considerably more complex than the laws of similitude and exchange used to derive measures in physics, and certainly it is of a wholly different, nonmeasurement character. Second, because of the lack of structure among psychophysical measures, there is considerable uncertainty about what units these inferred parameters should have. In many cases, they are defined in such a way as to be dimensionless, which differs considerably from physics. Third, although probability is an extensive measure of sorts, it is a somewhat unusual one; moreover, it would be amazing if the different psychophysical measures—for loudness, brightness, subjective length, etc.—should all be derived from the same extensive measure, response probability, whereas physics requires six extensive attributes.

All in all, the analogy to the structure of physical dimensions seems stretched all out of shape, and I doubt that it will prove useful in guiding future work.

### *Is Psychophysics the Study of a Measuring Device?*

Having tentatively rejected the notion that psychophysical measurement is comparable to either fundamental or derived physical measurement, we are left with the possible analogy to a measuring device. The version that I shall urge is that the sensory part of the organism is a mechanism whose

function is to recast all sorts of physical inputs into a common measure with which the central nervous system can deal. In this view, the task of psychophysics is to discover the nature of that recoding, to provide ways to ascertain the particular calibration of any given individual, and to understand how these measurements are processed by the organism so as to lead to responses. The various kinds of observed psychophysical data—magnitude estimates, comparisons of signals, detection, reaction time, etc.—should assist us in inferring exactly what that recoding and processing is.

According to this view, we must discover answers to two quite different questions. The first is the nature of the recoding from physical stimuli into some suitable, internal measure. The model for this should take into account the stimuli and the organism, but should be completely independent of the tasks we set the organism, the payoffs, and non-sensory variations in the experimental design. The second question concerns the sorts of processing that are carried out on these measures as the subject attempts to respond to the questions we ask of him. Here the model presupposes the theory of sensory transduction, and it varies with the experiment and the question asked, for surely the sensory information is not processed in the same way when we ask for a magnitude estimate as when we ask for a fast detection.

Consider the first problem, the nature of the transduction. Although most physical measuring devices transform an attribute to be measured into a measurement of length, that is obviously not the mode of operation in the nervous system. Some of the psychophysical theories designed to deal with confusable stimuli have postulated rather abstract recodings. Fechner supposed it to be the transformation that renders just noticeable differences equal; Thurstone postulated it to be one that renders the randomness of the representation normally distributed; the theory of signal detectability says it is the likelihood ratio of a hypothetical observation;<sup>6</sup> and the threshold theories assume a discrete set of ordered, abstract states with some sort of probability distribution over them. Each of these ideas has one or more drawbacks, among them the fact that some are untestable, some are

<sup>6</sup> The postulated normality of the distribution of likelihood ratio is entirely secondary in the theory of signal detectability.

limited to a very few experimental designs, and all lack neurophysiological support.

Recently, several psychophysicists (Creelman, 1964; Fitzhugh, 1958; Green & Luce, 1967, 1971, in press; Grice, 1968; Kohfeld, 1969; Luce, 1966; Luce & Green, 1970, 1972; McGill, 1963, 1967; McGill & Goldberg, 1968; Siebert, 1965, 1968, 1970) have begun to take seriously the neurophysiological evidence that signal intensity is encoded as pulse rates in the peripheral neural system (Kiang, 1965, 1968; Perkel & Bullock, 1968; Rose, Brugge, Anderson, & Hind, 1967). Here, then, is a possible convenient measure in terms of which all information about sensory intensity could be encoded—time. As an idealization—although neither an essential one nor, to judge by the data of Rose et al., a bad one—the pulse trains resulting from a constant intensity signal are assumed to be Poisson; that is, the times between successive pulses are independent random variables with an exponential distribution. The expected value of these interarrival times, which is the reciprocal of the pulse rate, has been shown to decrease systematically with intensity. This means that each interarrival time provides the nervous system with an independent estimate of the signal intensity producing the pulse train. Of course, the transformation from signal intensity to the Poisson intensity parameter is, in general, nonlinear and may be expected to vary from person to person and from ear to ear or eye to eye within a person.

Since the processing of recoded information leading to responses is almost certainly not carried out at the periphery, these neurophysiological data are not necessarily relevant to our problem of developing a suitable psychophysical theory. A priori, it is entirely possible that the representation has been altered considerably by the time it arrives at the central processing location. We assume not. As a working hypothesis, we suppose that signal intensity is represented at the processing center as a Poisson pulse train; however, we do not assume that the peripheral neurophysiological estimates of rates necessarily apply to the central process. In a sense, then, this type of theory is just as abstract as those I mentioned above; however, it has the advantages of paralleling the peripheral data, of being independent of the nonsensory aspects of the experiment, and of applying to all experimental designs in which the signal is used.

### *Some Consequences of Neural Pulse Models*

Given our working hypothesis about the sensory recoding, the problem is to devise plausible decision rules that are able to account for all psychophysical results. Clearly, the working through of a program of such scope will require the efforts of a number of people over some years. Here I simply cite a few encouraging results that David M. Green of the University of California, San Diego, and I have obtained during the past two years.

A decision rule operating on a pulse train can focus either on the number of pulses that occur in a fixed time or on the time required for a fixed number of pulses to occur. Each ratio provides an estimate of the Poisson parameter. The former has the advantage of taking a known amount of time, but the sample size, and therefore the statistical quality of the estimate, varies; the latter has a predetermined sample size, but the time needed for the estimate is variable. Models that assume a count of the pulses occurring in a fixed time are known, naturally enough, as counting models (McGill, 1967). Those in which a fixed number of interarrival times are accumulated are known as timing models (Luce & Green, 1972).

A first question is whether we can dismiss either class of rules. The answer seems to be no—apparently both modes of operation can occur. The evidence for this statement comes from a yes-no, auditory detection experiment using a response-terminated signal and a response deadline (Green & Luce, in press). Responses occurring prior to the signal onset or after the deadline were fined; those in between were paid off for accuracy. We varied both the deadline for a fixed symmetric payoff matrix and the payoff matrix for a fixed deadline. It is not difficult to show that the counting model implies that the response times should depend neither on the signal presented nor on the response made. Further, the receiver operating characteristic (ROC curve) in a normal-normal probability plot should be approximately linear and have a slope less than one. The timing model implies that variation of the deadline should lead to a linear relation between the mean response time to noise and the mean response time to signal, and variation of the payoffs should lead to an approximately linear ROC plot. Moreover, both functions should have the same slope and it should be greater than one—which is virtually unheard of

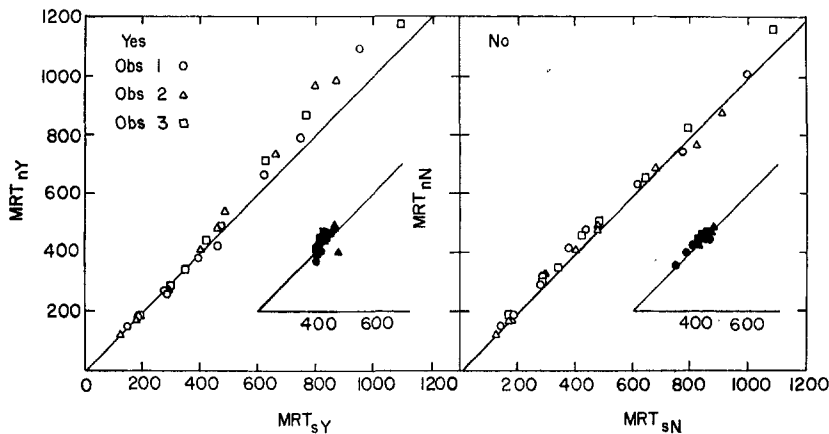


FIG. 1. Mean reaction time plots using the ordinary deadline procedure. (The first subscript is the stimulus condition—noise [n] or signal [s]. The second is the response condition—Yes [Y] or No [N]. The open points were generated by varying the deadline for a fixed symmetric payoff matrix; the solid ones, in the insert, by varying the payoff matrix for a fixed 600-millisecond deadline.)

for empirical ROC plots constructed from signals of fixed duration. The mean response times for the conventional deadline procedure are shown in Figure 1, and the corresponding ROC curve is shown in Figure 2. Clearly, of these two models, the counting one is correct. But now consider an apparently minor variant of the experiment in which the deadline applies only to signal trials, not to noise ones. The comparable results are shown in Figures 3 and 4. Clearly, the timing model is now favored.<sup>7</sup>

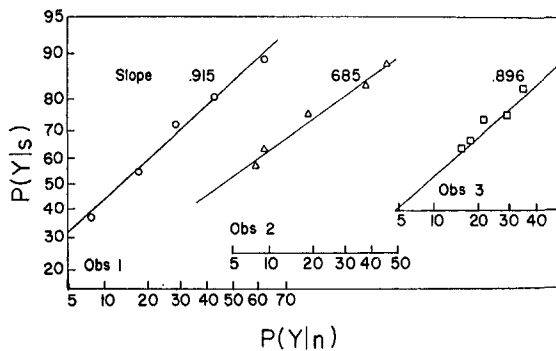


FIG. 2. ROC curves, on double probability coordinates, obtained from a fixed 600-millisecond deadline by varying the payoff matrix.

<sup>7</sup> The close agreement of the mean reaction time and ROC slopes, which according to the theory should be identical, is probably more apparent than real. Without going into details, the difficulty is that the slope of the ROC curve is very sensitive to small discrepancies in the probabilities near 0 and 1, and the approximation of the gamma distribution by the normal introduces just such discrepancies.

We do not yet have any certain idea as to the conditions that elicit counting rather than timing behavior. My personal guess is that in its ordinary environment an organism simply monitors short, overlapping sequences of interarrival times, attending to marked changes in the average time as evidence for a change in the intensity of that signal. Counting is the preferred mode of operation only in the peculiar environment of those experiments having fixed, well-marked observation periods. Even then, as we have just seen, the counting behavior can be replaced by timing behavior. Presumably the reason that the counting mode is available, even though it cannot be very useful in the natural environment, is that both modes of operation require both counting and timing mechanisms in the nervous system. Therefore, models of both types are going to have to be worked out in detail. Since Green and I have emphasized timing ones, I will limit myself to some illustrative results from these.

In a magnitude estimation experiment, the subject is asked to report numerically the level of intensity of each signal. In terms of a pulse-rate model, this can be interpreted as asking him to estimate and report the rate. If so, in the timing mode of operation, he collects a fixed number of interarrival times (the number being limited by signal duration or by some form of memory), averages them, and reports the reciprocal.<sup>8</sup> Thus, Stev-

<sup>8</sup> If the experimenter has set an arbitrary unit, then the reciprocal has to be multiplied by a positive constant.

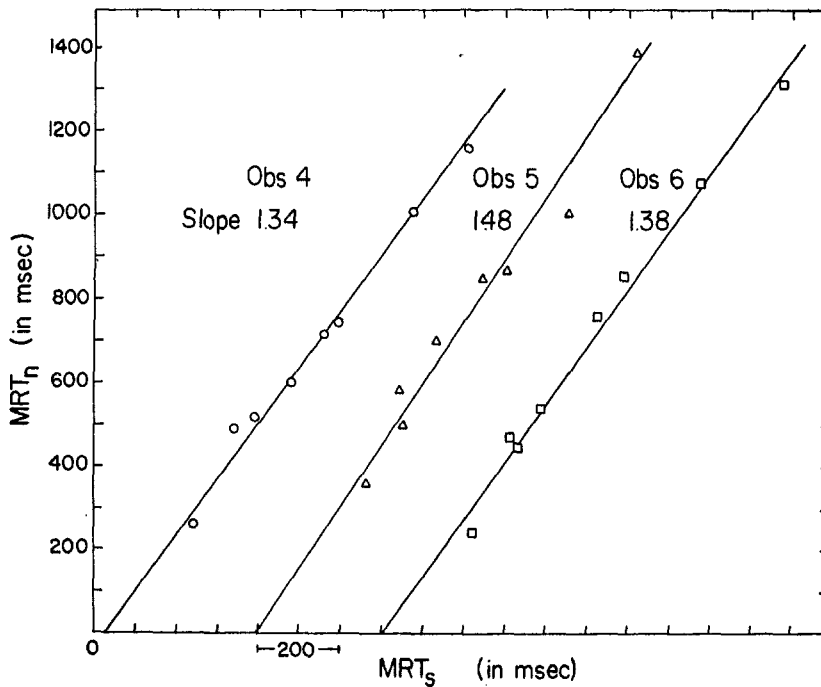


FIG. 3. Mean reaction time to noise versus that to signal (combined over responses) using a variable deadline only on signal trials.

ens' (1957) data (and numerous later publications by him and others) say that the Poisson parameter is approximately a power function of intensity; moreover, the exponent varies over a fairly wide range with both modality and subject. Stevens (1970) summarized data in which physiological

measures appear to grow with intensity in the same way as magnitude estimates. The timing theory also predicts that the frequency distribution of these estimates at any one level of intensity has a high tail—of the form  $t^k$  rather than  $e^{-at}$ —which result was noted empirically by Luce and Mo

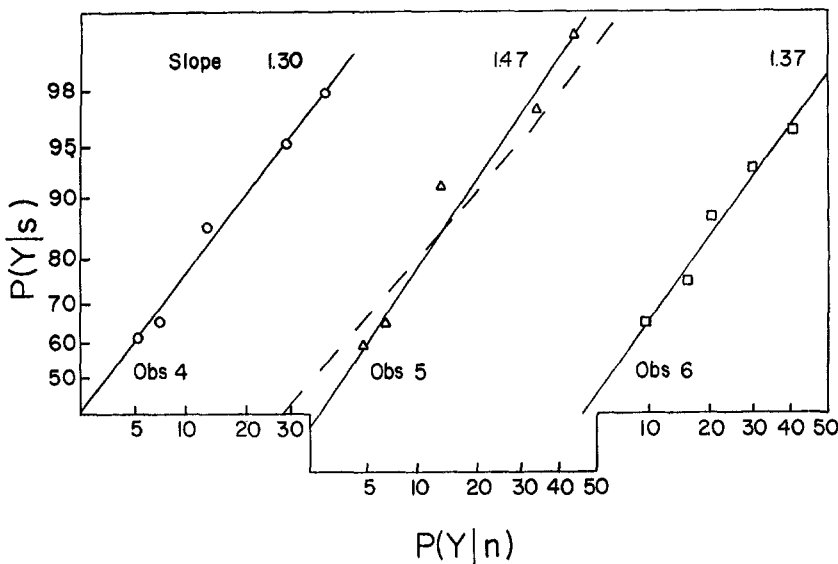


FIG. 4. ROC curves, on double probability coordinates, obtained from a fixed 600-millisecond deadline on signal trials by varying the payoff matrix. (The dashed line is the corresponding gamma distribution, which shows how sensitive the slope is to minor discrepancies in the tails.)



(1965). Our estimate of the number of interarrival times entering into calculations of magnitude responses is not very solidly based at the present time, but it appears to be in the range from 10 to 25.

A timing analysis of the simple reaction-time, random foreperiod experiment is easily described, although not so easily worked out. We assume that each interarrival time is compared with a criterion, and the first one less than the criterion is interpreted as evidence for the onset of the signal. Among other things, this leads to a prediction which indirectly favors the above power function conclusion, namely, that for signals of relatively long duration, the mean reaction time should be of the form  $MRT = \bar{r} + B/ME$ , where  $\bar{r}$  and  $B$  are positive constants and  $ME$  is the mean magnitude estimate of the same signal.<sup>9</sup> Vaughan, Costa, and Gilden (1966) reported that this holds for brightness; however, Mansfield (1970) showed that it breaks down for durations below 10 milliseconds. By choosing  $\bar{r}$  to be 5 milliseconds less than the mean reaction time corresponding to the most intense signal, the three sets of loudness data shown in Figure 5 support the same hypothesis.<sup>10</sup> In order for a counting model to handle these data, we would have to make a special ad hoc assumption about the observation interval changing with signal intensity.

Turning to discrimination experiments, suffice it to say that the timing models resemble non-constant variance Thurstonian models in the sense that the total time to collect a fixed number of interarrival times is a random variable that is compared with another similar time. The major difference is that the distribution of these random variables is gamma rather than normal. A necessary and sufficient condition for Weber's law to hold in this model is that the Poisson parameter be a power function of signal intensity. Moreover, the well-known increase in the Weber fraction at low intensity has a completely natural explanation in

<sup>9</sup> Of course, the reaction-time theory actually predicts a linear relation between mean reaction time and the reciprocal of the Poisson parameter. Combining this with the theory for the magnitude experiment leads to the prediction stated.

<sup>10</sup> McGill (1963), using a different estimate for  $\bar{r}$ , concluded otherwise; the slope, but not the linearity, of the plot is markedly affected by changes of a few milliseconds in the value selected for  $\bar{r}$ .

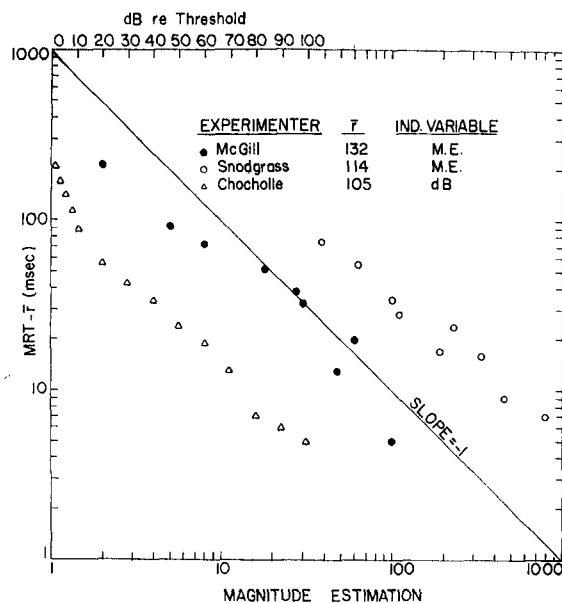


FIG. 5. Plots of  $MRT - \bar{r}$  versus  $ME$  of loudness of 1000 Hertz tones from three experiments. (For Chocholle's [1940] data, it is assumed that  $ME = \alpha I^{0.8}$ . In each case,  $\bar{r}$  was chosen to be 5 milliseconds less than the minimum mean reaction time; this is different from McGill's [1963] choice, and it alters radically his conclusion. The data of J. G. Snodgrass are unpublished. Reprinted from an article by Luce and Green from the January 1972 *Psychological Review*. Copyrighted by the American Psychological Association, Inc., 1972.)

the timing model, namely, that the number of interarrival times that can be expected to be observed during a fixed signal interval decreases with signal intensity, and so the precision with which the signal intensity is defined must, of necessity, be reduced. The constancy of the Weber fraction beyond a certain intensity suggests either a limit to the number of interarrival times that can be collected and stored or an asymptote to the function relating the Poisson parameter to intensity. We do not yet know which is the better explanation.

### Concluding Remarks

My main contention is that existing psychophysical measures do not exhibit one of the major invariances required if they are to have a structure at all comparable to physical measures. I conclude that it is more appropriate to think of sensory systems simply as devices that recode the intensity of various types of physical signals into a common neural measure, namely, time between pulses or, equivalently, pulse rate. Assuming that such a

view is correct, what consequence does it have for the standardization of psychophysical scales such as is typified by the sone scale of loudness? I interpret such standard scales simply as statements about the approximate calibration of a typical person as a measuring device. When it is said that loudness is a power function with an exponent of .3, this means that, to some approximation, the measuring device effects a power transformation from sound intensity to a time measure in the nervous system and that the average exponent over some population of people is .3. The attempt to attach units of loudness to such a standard scale is, I believe, completely misguided and is based on the mistaken notion that one is discovering fundamental measures comparable to mass, energy, and the like. The value in having standard scales, which is very great indeed, depends not on their measurement theoretic character, but on the fact that they describe how the typical person transduces physical intensity into the measure used by his central nervous system. Knowing this permits us to approach more rationally social decisions about such matters as acceptable and unacceptable noise levels.

In closing, let me explicitly cite three things that were not said and which I do not intend to imply. First, by saying that psychophysics is concerned with the workings and calibration of measuring devices and not with the discovery of new fundamental measures, I do not intend to denigrate psychophysics; I have emphasized the point only because it seems useful to be quite clear as to what one is about. Second, I do not imply that my conclusions apply beyond psychophysical measurement. I am not at all sure whether other branches of psychology ultimately will develop fundamental measures distinct from those already known to physics. A number of apparently ordinal concepts, among them hunger, anxiety, fear, drive, have so far resisted any systematic measurement analysis. Perhaps when we gain more insight into them, they will assume a status as fundamental attributes. And third, my conclusions in no way reflect adversely upon the usefulness of ordinal theories of measurement in psychology, in particular in psychophysics. Theories such as additive conjoint measurement, which formulate ordinally certain forms of independence and show how to construct economical representations of it, provide initial steps toward simple theories of interaction.

All that I urge is that we not impute to such scales more structure than we have actually demonstrated.

#### REFERENCES

- BOCK, R. D., & JONES, L. V. *The measurement and prediction of judgment and choice*. San Francisco: Holden-Day, 1968.
- BRADLEY, R. A., & TERRY, M. E. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 1952, 39, 324-345.
- CAMPBELL, B. A., & MASTERTSON, F. A. Psychophysics of punishment. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior*. New York: Appleton-Century-Crofts, 1969.
- CHOCHOLLE, R. Variations des temps de réaction auditifs en fonction de l'intensité à diverses fréquences. *L'Année Psychologique*, 1940, 41, 5-124.
- COOMBS, C. H. *Theory of data*. New York: Wiley, 1964.
- CREELMAN, C. D. Human discrimination of auditory duration. *Journal of the Acoustical Society of America*, 1961, 34, 582-593. (Reprinted in J. A. Swets [Ed.], *Signal detection and recognition by human observers*. New York: Wiley, 1964.)
- DEBREU, G. Topological methods in cardinal utility theory. In K. J. Arrow, S. Karlin, & P. Suppes (Eds.), *Mathematical methods in the social sciences, 1959*. Stanford: Stanford University Press, 1960.
- EISLER, H. A general differential equation in psychophysics: Derivation and empirical test. *Scandinavian Journal of Psychology*, 1963, 4, 265-272.
- FALMAGNE, J. C. The generalized Fechner problem and discrimination. *Journal of Mathematical Psychology*, 1971, 8, 22-43.
- FECHNER, G. T. *Elemente der Psychophysik*. Leipzig: Breitkopf und Hartel, 1860. (Trans. by H. E. Adler, *Elements of psychophysics*. New York: Holt, 1966.)
- FITZHUGH, R. A statistical analyzer for optic nerve messages. *Journal of General Physiology*, 1958, 41, 675-692.
- GREEN, D. M., & LUCE, R. D. Detection of auditory signals presented at random times. *Perception & Psychophysics*, 1967, 2, 441-449.
- GREEN, D. M., & LUCE, R. D. Detection of auditory signals presented at random times, III. *Perception & Psychophysics*, 1971, 9, 257-268.
- GREEN, D. M., & LUCE, R. D. Speed-accuracy trade-off in auditory detection. In S. Kornblum (Ed.), *Attention and performance*. Vol. 4. New York: Academic Press, 1972, in press.
- GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.
- GRICE, G. R. Stimulus intensity and response evocation. *Psychological Review*, 1968, 75, 359-373.
- HELMHOLTZ, H. V. Zählen und Messen erkenntnis-theoretisch betrachtet. *Philosophische Aufsätze Eduard Zeller gewidmet*. Leipzig, 1887. (Trans. by C. L. Bryan, *Counting and measuring*. Princeton, N.J.: Van Nostrand, 1930.)
- HÖLDER, O. Die Axiome der Quantität und die Lehre vom Mass. *Berichte der Sächsischen Gesellschaft der Wissenschaften, Leipzig, mathematisch-physikalisch Klasse*, 1901, 53, 1-64.

- KIANG, N. Y-S. *Discharge patterns of single fibers in the cat's auditory nerve*. Cambridge: M.I.T. Press, 1965.
- KIANG, N. Y-S. A survey of recent developments in the study of auditory physiology. *Annals of Otolaryngology and Rhinology*, 1968, 77, 656-675.
- KOHFELD, D. L. Effects of the intensity of auditory and visual ready signals on simple reaction times. *Journal of Experimental Psychology*, 1969, 82, 88-95.
- KRANTZ, D. H. A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology*, 1972, in press.
- KRANTZ, D. H., LUCE, R. D., SUPPES, P., & TVERSKY, A. *Foundations of measurement*. Vol. 1. New York: Academic Press, 1971.
- LEVELT, W. J. M., RIEMERSMA, J. B., & BUNT, A. A. Binaural additivity of loudness. *British Journal of Mathematical and Statistical Psychology*, 1971, in press.
- LEVINE, M. V. Transformations that render curves parallel. *Journal of Mathematical Psychology*, 1970, 7, 410-443.
- LEVINE, M. V. Transformations that render curves similar and linearly related random variables. *Journal of Mathematical Psychology*, 1972, in press.
- LUCE, R. D. *Individual choice behavior: A theoretical analysis*. New York: Wiley, 1959.
- LUCE, R. D. A threshold theory for simple detection experiments. *Psychological Review*, 1963, 70, 61-79.
- LUCE, R. D. A model for detection in temporally unstructured experiments with a Poisson distribution of signal presentations. *Journal of Mathematical Psychology*, 1966, 3, 48-64.
- LUCE, R. D. Review of C. W. Savage's *The measurement of sensation*. *Science*, 1971, 171, 165-166.
- LUCE, R. D., & EDWARDS, W. The derivation of subjective scales from just noticeable differences. *Psychological Review*, 1958, 65, 222-237.
- LUCE, R. D., & GREEN, D. M. Detection of auditory signals presented at random times, II. *Perception & Psychophysics*, 1970, 7, 1-14.
- LUCE, R. D., & GREEN, D. M. A neural timing theory for response times and the psychophysics of intensity. *Psychological Review*, 1972, 79, 14-57.
- LUCE, R. D., & MO, S. S. Magnitude estimation of heaviness and loudness by individual subjects: A test of a probabilistic response theory. *British Journal of Mathematical and Statistical Psychology*, 1965, 18, 159-174.
- LUCE, R. D., & TUKEY, J. Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1964, 1, 1-27.
- MANSFIELD, R. J. W. Intensity relations in vision: Analysis and synthesis of a nonlinear sensory system. Unpublished doctoral dissertation, Harvard University, 1970.
- MCGILL, W. J. Stochastic latency mechanisms. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. 1. New York: Wiley, 1963.
- MCGILL, W. J. Neural counting mechanisms and energy detection in audition. *Journal of Mathematical Psychology*, 1967, 4, 351-376.
- MCGILL, W. J., & GOLDBERG, J. P. Pure-tone intensity discrimination and energy detection. *Journal of the Acoustical Society of America*, 1968, 44, 571-581.
- PERKEL, D. H., & BULLOCK, T. H. Neural coding. *Bulletin, Neurosciences Research Program*, 1968, 6, 221-348.
- PFANZAGL, J. A general theory of measurement—applications to utility. *Naval Research Logistics Quarterly*, 1959, 6, 283-294.
- PFANZAGL, J. *Theory of measurement*. New York: Wiley, 1968.
- ROSE, J. E., BRUGGE, J. F., ANDERSON, D. J., & HIND, J. E. Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. *Journal of Neurophysiology*, 1967, 30, 769-793.
- SAVAGE, C. W. *The measurement of sensation*. Berkeley: University of California Press, 1970.
- SIEPARD, R. N. Metric structures in ordinal data. *Journal of Mathematical Psychology*, 1966, 3, 287-315.
- SIEBERT, W. M. Some implications of the stochastic behavior of primary auditory neurons. *Kybernetik*, 1965, 2, 206-215.
- SIEBERT, W. M. Stimulus transformations in the peripheral auditory system. In P. A. Kohlers & M. Eden (Eds.), *Recognizing patterns*. Cambridge: M.I.T. Press, 1968.
- SIEBERT, W. M. Frequency discrimination in the auditory system: Place or periodicity mechanisms? *Proceedings of the IEEE*, 1970, 58, 723-730.
- STEVENS, J. C., & HALL, J. W. Brightness and loudness as functions of stimulus duration. *Perception & Psychophysics*, 1966, 1, 319-327.
- STEVENS, J. C., & STEVENS, S. S. Brightness function: Effects of adaptation. *Journal of the Optical Society of America*, 1963, 53, 375-385.
- STEVENS, S. S. On the psychophysical law. *Psychological Review*, 1957, 64, 153-181.
- STEVENS, S. S. Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *Journal of Experimental Psychology*, 1959, 57, 201-209.
- STEVENS, S. S. Duration, luminance, and the brightness exponent. *Perception & Psychophysics*, 1966, 1, 96-100. (a)
- STEVENS, S. S. Matching functions between loudness and ten other continua. *Perception & Psychophysics*, 1966, 1, 5-8. (b)
- STEVENS, S. S. Neural events and the psychophysical law. *Science*, 1970, 170, 1043-1050.
- STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, 1957, 54, 377-411.
- SUPPES, P., & ZINNES, J. L. Basic measurement theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. 1. New York: Wiley, 1963.
- TANNER, W. P., JR., & SWETS, J. A. A decision-making theory of visual detection. *Psychological Review*, 1954, 61, 401-409.
- THURSTONE, L. L. *The measurement of values*. Chicago: University of Chicago Press, 1947.
- VAUGHAN, H. G., COSTA, L. D., & GILDEN, L. The functional relation of visual evoked response and reaction time to stimulus intensity. *Vision Research*, 1966, 6, 645-656.