VOL. 66, NO. 2 MARCH, 1959

THE PSYCHOLOGICAL REVIEW

ON THE POSSIBLE PSYCHOPHYSICAL LAWS [1]

R. DUNCAN LUCE

Harvard University

This paper is concerned with the century-old effort to determine the functional relations that hold between subjective continua and the physical continua that are presumed to underlie them. The first, and easily the most influential, attempt to specify the possible relations was made by Fechner. It rests upon empirical knowledge of how discrimination varies with intensity along the physical continuum and upon the assumption that jnd's are subjectively equal throughout the continuum. When, for example, discrimination is proportional to intensity (Weber's law), Fechner claimed that the equal-jnd assumption leads to a logarithmic relation (Fechner's law).

This idea has always been subject to controversy, but recent attacks upon it have been particularly severe. At the theoretical level, Luce and Edwards (1958) have pointed out that Fechner's mathematical reasoning was not sound. Among other things, his assumption is not sufficient to generate an interval scale. By recasting his problem somewhat—essentially by replacing the equal-jnd assumption with the somewhat stronger condition that "equally often noticed differences are equal, except when always or never noticed"—they were able to show that an interval scale results, and to present a mathematical expression for it. Their work has no practical import when Weber's law, or its linear generalization $\Delta x = ax + b$, is true, because the logarithm is still the solution, but their jnd scale differs from Fechner's integral when Weber's law is replaced by some other function relating stimulus jnd's to intensity.

At the empirical level, Stevens (1956, 1957) has argued that jnd's are unequal in subjective size on intensive, or what he calls prothetic, continua—a contention supported by considerable data—and that the relation between the subjective and physical continua is the power function αx^{β}, not the logarithm. Using such "direct" methods as magnitude estimation and ratio production, he and others (Stevens: 1956, 1957; Stevens & Galanter, 1957) have accumulated considerable evidence to but-

[1] This work has been supported in part by Grant M-2293 from the National Institute of Mental Health and in part by Grant NSF-G 5544 from the National Science Foundation.

Ward Edwards, E. H. Galanter, Frederick Mosteller, Frank Restle, S. S. Stevens, and Warren Torgerson have kindly given me their thoughtful comments on drafts of this paper, many of which are incorporated into this version. I am particularly indebted to S. S. Stevens for his very detailed substantive and stylistic criticisms of the last two drafts.

tress the empirical generality of the power function. Were it not for the fact that some psychophysicists are uneasy about these methods, which seem to rest heavily upon our experience with the number system, the point would seem to be established. In an effort to bypass these objections, Stevens (1959) has recently had subjects match values between pairs of continua, and he finds that the resulting relations are power functions whose exponents can be predicted from the magnitude scales of the separate variables. Thus, although much remains to be learned about the "direct" methods of scaling, the resulting power functions appear to summarize an interesting body of data.

Given these empirical results, one is challenged to develop a suitable formal theory from which they can be shown to follow. There can be little doubt that, as a starting point, certain commonly made assumptions are inappropriate: equality of jnd's, equally often noticed differences, and Thurstone's equal variance assumption. Since, however, differences stand in the same —logarithmic—relation to ratios as Fechner's law does to the power function, a reasonable starting point might seem to be the assumption that the subjective ratio of stimuli one jnd apart is a constant independent of the stimulus intensity. Obvious as the procedure may seem, in my opinion it will not do. Although generations of psychologists have managed to convince themselves that the equal-jnd assumption is plausible, if not obvious, it is not and never has been particularly compelling; and in this respect, an equal-ratio assumption is not much different. This is not to deny that subjective continua may have the equal-ratio property—they must if the power law is correct and Weber's law holds—but rather to argue that such an assumption is too special to be acceptable as a basic axiom in a deductive theory.

Elsewhere (Luce, in press), I have suggested another approach. An axiom, or possible law, of wide applicability in the study of choice behavior, may be taken in conjunction with the linear generalization of Weber's law to demonstrate the existence of a scale that is a power function of the physical continuum. Although that theory leads to what appears to be the correct form, it is open to two criticisms. First, the exponent predicted from discrimination data is at least an order of magnitude larger than that obtained by direct scaling methods. Second, the theory is based upon assumptions about discriminability, and these are not obviously relevant to a scale determined by another method. Scales of apparent magnitude may be related to jnd scales, but it would be unwise to take it for granted that they are.

The purpose of this paper is to outline still another approach to the problem, one that is not subject to the last criticism. The results have applicability far beyond the bounds of psychophysics, for they concern the general question of the relation between measurement and substantive theories.

TYPES OF SCALES

Although familiarity may by now have dulled our sense of its importance, Stevens' (1946, 1951) stress upon the transformation groups that leave certain specified scale properties invariant must, I think, be considered one of the more striking contributions to the discussion of measurement in the past few decades. Prior to his work, most writers had put extreme emphasis upon the property of "additivity," which is a characteristic of much physical measurement (Cohen & Nagel, 1934). It was held that this property is fundamental to scientific measurement and, indeed, the term "fundamental measurement" was applied only to these scales. This contention, however, puts the nonphysical sciences in a most peculiar fix. Since no one has yet discovered an "additive" psychological variable, it would seem that psychology can have no fundamental measures of its own. This conclusion might be acceptable if we could define psychological measures in terms of the fundamental physical scales, i.e., as "derived" scales, but few of the things we want to measure seem to be definable in this way. So either rigorous psychological measurement must be considered impossible or additive empirical operations must not be considered essential to measurement. What is important is not additivity itself, but the fact that, when it is coupled with other plausible assumptions, it sharply restricts the class of transformations that may be applied to the resulting scale. Specifically, it makes the scale unique except for multiplication by positive constants, i.e., changes of unit. Additivity is not the only property that an assignment of numbers to objects or events may have which sharply limits the admissible transformations. Some of these other properties appear applicable and relevant to psychological variables, and so in this sense psychological measurement appears to be possible.

By a *theory of measurement,* I shall mean the following. One or more operations and relations are specified over a set of objects or events (the variable), and they are characterized by a number of empirically testable assumptions. In addition, it must be possible to assign numbers to the objects and to identify numerical operations and relations with the empirical operations and relations in such a way that the numerical operations represent (are isomorphic to) the empirical ones. In other words, we have a measurement theory whenever (*a*) we have a system of rules for assigning numerical values to objects that are interrelated by assumptions about certain empirical operations involving them, and (*b*) these rules let us set up an isomorphic relation between some properties of the number system and some aspects of the empirical operations.

One of the simplest examples of a theory of measurement is a finite set (of goods) ordered by a binary (preference) relation P that is assumed to be antisymmetric and transitive. A scale u can be assigned to the set in such a manner that it represents P in the sense that xPy if and only if $u(x) > u(y)$.

By the *scale type,* I shall mean the group of transformations that result in other isomorphic representations of the measurement theory. In the preceding example any strictly monotonic increasing transformation will do, and scales of this type are known as ordinal. Any transformation chosen from the scale type will be said to be an *admissible transformation.*

It should be re-emphasized that quite divergent measurement theories may lead to the same scale type. For example, Case V of Thurstone's law of comparative judgment (1927) and the von Neumann-Morgenstern utility axioms (1947) both result in interval scales (of something), yet the basic terms and assumptions involved are totally different, even though both theories can be applied to the same subject matter. Of course, the resulting interval scales may not be linearly related, for they may be measuring different things.

A measurement theory may be contrasted with what I shall call a *substantive theory.* The former involves operations and assumptions only about a single class of objects which is treated as a unitary variable, whereas the latter involves relations among two or more variables. In practice, substantive theories are usually stated in terms of functional relations among the scales that result from the several measurement theories for the variables involved.

For a number of purposes, the scale type is much more crucial than the details of the measurement theory from which the scale is derived. For example, much attention has been paid to the limitations that the scale type places upon the statistics one may sensibly employ. If the interpretation of a particular statistic or statistical test is altered when admissible scale transformations are applied, then our substantive conclusions will depend upon which arbitrary representation of the scale we have used in making our calculations. Most scientists, when they understand the problem, feel that they should shun such statistics and rely only upon those that exhibit the appropriate invariances for the scale type at hand. Both the geometric and arithmetic means are legitimate in this sense for ratio scales (unit arbitrary), only the latter is legitimate for interval scales (unit and zero arbitrary), and neither for ordinal scales. For fuller discussions, see Stevens: 1946, 1951, 1955; for a somewhat less strict interpretation of the conclusions, see Mosteller, 1958.

A second place where the transformation group imposes limitations is in the construction of substantive theories. These limitations seem to have received far less attention than the statistical questions, even though they are undoubtedly more fundamental. The remainder of the paper will attempt to formulate the relation between scale types and functional laws, and to answer the question what psychophysical laws are possible. As already pointed out, these issues have scientific relevance beyond psychophysics.

A PRINCIPLE OF THEORY CONSTRUCTION

In physics one finds at least two classes of basic assumptions: specific empirical laws, such as the universal law of gravitation or Ohm's law, and a priori principles of theory construction, such as the requirement that the laws of mechanics should be invariant under uniform translations and rotations of the coordinate system. Other laws, such as the conservation of energy, seem to have changed from the empirical to the a priori category during the development of physics. In psychology more stress has been put on the discovery of empirical laws than on the formulation of guiding principles, and the search for empirical relations tends to be pursued without the benefit of explicit statements about what is and is not an acceptable theory.[2] Since such principles have been used effectively in physics to limit the possible physical laws, one wonders whether something similar may not be possible in psychology.

Without such principles, practically any relation is a priori possible, and the correct one is difficult to pin down by empirical means because of the ever present errors of observation. The error problem is particularly acute in the behavioral sciences. On the other hand, if a priori consideration about what constitutes an acceptable theory limits us to some rather small set of possible laws, then fairly crude obser-

[2] Two attempts to introduce and use such statements in behavioral problems are the combining of classes condition in stochastic learning theory (Bush, Mosteller, & Thompson, 1954) and some work on the form of the utility function for money which is based upon the demand that certain game theory solutions should remain unchanged when a constant sum of money is added to all the payoffs (Kemeny & Thompson, 1957). In neither case do the conditions seem particularly compelling.

vations may sometimes suffice to decide which law actually obtains.

The principle to be suggested appears to be a generalization of one used in physics. It may be stated as follows.

> A substantive theory relating two or more variables and the measurement theories for these variables should be such that:
>
> 1. (*Consistency of substantive and measurement theories*) Admissible transformations of one or more of the independent variables shall lead, via the substantive theory, only to admissible transformations of the dependent variables.
>
> 2. (*Invariance of the substantive theory*) Except for the numerical values of parameters that reflect the effect on the dependent variables of admissible transformations of the independent variables, the mathematical structure of the substantive theory shall be independent of admissible transformations of the independent variables.

In this principle, and in what follows, the terms independent and dependent variables are used only to distinguish the variables to which arbitrary, admissible transformations are imposed from those for which the transformations are determined by the substantive theory. As will be seen, in some cases the labeling is truly arbitrary in the sense that the substantive theory can be written so that any variable appears either in the dependent or independent role, but in other cases there is a true asymmetry in the sense that some variables must be dependent and others independent if any substantive theory relates them at all.

One can hardly question the consistency part of the principle. If an admissible transformation of an independent variable leads to an inadmissible transformation of a dependent variable, then one is simply saying that the strictures imposed by the measurement theories are incompatible with those imposed by the substantive theory. Such a logical inconsistency must, I think, be interpreted as meaning that something is amiss in the total theoretical structure.

The invariance part is more subtle and controversial. It asserts that we should be able to state the substantive laws of the field without reference to the particular scales that are used to measure the variables. For example, we want to be able to say that Ohm's law states that voltage is proportional to the product of resistance and current without specifying the units that are used to measure voltage, resistance, or current. Put another way, we do not want to have one law when one set of units is used and another when a different set of units is used. Although this seems plausible, there are examples from physics that can be viewed as a particular sort of violation of Part 2; however, let us postpone the discussion of these until some consequences of the principle as stated have been derived.

The meaning of the principle may be clarified by examples that violate it. Suppose it is claimed that two ratio scales are related by a logarithmic law. An admissible transformation of the independent variable x is multiplication by a positive constant k, i.e., a change of unit. However, the fact that $\log kx = \log k + \log x$ means that an inadmissible transformation, namely, a change of zero, is effected on the dependent variable. Hence, the logarithm fails to meet the consistency requirement. Next, consider an exponential law, then the transformation leads to $e^{kx} = (e^x)^k$. This can be viewed either as a violation of consistency or of invariance. If the law is exponential, then the dependent variable is raised to a power, which is inconsistent with its being a ratio scale. Alternatively, the dependent variable may be taken to be a ratio scale, but then the law is not invariant because it is an exponential raised to a power that depends upon the unit of the independent variable.

AN APPLICATION OF THE PRINCIPLE

Most of the physical measures entering into psychophysics are idealized in physical theories in such a way that they form either ratio or interval scales. Mass, length, pressure, and time durations are measured on ratio scales, and physical time (not time durations), ordinary temperature, and entropy are measured on interval scales. Of course, differences and derivatives of interval scale values constitute ratio scales.

Although most psychological scales in current use can at best be considered to be ordinal, those who have worked on psychological measurement theories have attempted to arrive at scales that are either ratio or interval, preferably the former. Examples: the equally often noticed difference assumption and the closely related Case V of Thurstone's "law of comparative judgment" lead to interval scales; Stevens has argued that magnitude estimation methods result in ratio scales (but no measurement theory has been offered in support of this plausible belief); and I have given sufficient conditions to derive a ratio scale from discrimination data. Our question here, however, is not how well psychologists have succeeded in perfecting scales of one type or another, but what a knowledge of scale types can tell us about the relations among scales.

In addition to these two common types of scales, there is some interest in what have been called logarithmic interval scales (Stevens, 1957). In this case the admissible transformations are multiplications by positive constants and raising to positive powers, i.e., kx^c, where $k > 0$ and $c > 0$. The name applied to this scale type reflects the fact that $\log x$ is an interval scale, since the transformed scale goes into $c \log x + \log k$. We will consider all combinations of ratio, interval, and logarithmic interval scales.

Because this topic is more general than psychophysics, I shall refer to the variables as independent and dependent rather than physical and psychological. Both variables will be assumed to form numerical continua having more than one point. Let $x \geq 0$ denote a typical value of the independent variable and $u(x) \geq 0$ the corresponding value of the dependent variable, where u is the unknown functional law relating them. Suppose, first, that both variables form ratio scales. If the unit of the independent variable is changed by multiplying all values by a positive constant k, then according to the principle stated above only an admissible transformation of the dependent variable, namely multiplication by a positive constant, should result and the form of the functional law should be unaffected. That is to say, the changed unit of the dependent variable may depend upon k, but it shall not depend upon x, so we denote it by $K(k)$. Casting this into mathematical terms, we obtain the functional equation

$$u(kx) = K(k)u(x)$$

where $k > 0$ and $K(k) > 0$.

Functional equations for the other cases are arrived at in a similar manner. They are summarized in Table 1.

The question is: What do these nine functional equations, each of which embodies the principle, imply about

TABLE 1

THE FUNCTIONAL EQUATIONS FOR THE LAWS SATISFYING THE PRINCIPLE OF THEORY CONSTRUCTION

Eq. No.	Scale Types		Functional Equation	Comments
	Independent Variable	Dependent Variable		
1	ratio	ratio	$u(kx)=K(k)u(x)$	$k>0,\ K(k)>0$
2	ratio	interval	$u(kx)=K(k)u(x)+C(k)$	$k>0,\ K(k)>0$
3	ratio	log interval	$u(kx)=K(k)u(x)^{C(k)}$	$k>0,\ K(k)>0,\ C(k)>0$
4	interval	ratio	$u(kx+c)=K(k,c)u(x)$	$k>0,\ K(k,c)>0$
5	interval	interval	$u(kx+c)=K(k,c)u(x)+C(k,c)$	$k>0,\ K(k,c)>0$
6	interval	log interval	$u(kx+c)=K(k,c)u(x)^{C(k,c)}$	$k>0,\ K(k,c)>0,\ C(k,c)>0$
7	log interval	ratio	$u(kx^c)=K(k,c)u(x)$	$k>0,\ c>0,\ K(k,c)>0$
8	log interval	interval	$u(kx^c)=K(k,c)u(x)+C(k,c)$	$k>0,\ c>0,\ K(k,c)>0$
9	log interval	log interval	$u(kx^c)=K(k,c)u(x)^{C(k,c)}$	$k>0,\ c>0,\ K(k,c)>0,\ C(k,c)>0$

the form of u? We shall limit our consideration to theories where u is a continuous, nonconstant function of x.

Theorem 1. If the independent and dependent continua are both ratio scales, then $u(x) = \alpha x^{\beta}$, where β is independent of the units of both variables.[3]

Proof. Set $x = 1$ in Equation 1, then $u(k) = K(k)u(1)$. Because u is nonconstant we may choose k so that $u(k) > 0$, and because $K(k) > 0$, it follows that $u(1) > 0$, so $K(k) = u(k)/u(1)$. Thus, Equation 1 becomes $u(kx) = u(k)u(x)/u(1)$. Let $v = \log[u/u(1)]$, then

$$\begin{aligned} v(kx) &= \log [u(kx)/u(1)] \\ &= \log \frac{u(k)u(x)}{u(1)u(1)} \\ &= \log [u(k)/u(1)] \\ &\qquad + \log [u(x)/u(1)] \\ &= v(k) + v(x) \end{aligned}$$

Since u is continuous, so is v, and it is well known that the only continuous solutions to the last functional equation are of the form

$$\begin{aligned} v(x) &= \beta \log x \\ &= \log x^{\beta} \end{aligned}$$

Thus,

$$\begin{aligned} u(x) &= \alpha e^{v(x)} \\ &= \alpha x^{\beta} \end{aligned}$$

where $\alpha = u(1)$.

We observe that since

$$u(kx) = \alpha k^{\beta} x^{\beta} = \alpha' x^{\beta}$$

β is independent of the unit of x, and it is clearly independent of the unit of u.

Theorem 2. If the independent continuum is a ratio scale and the depend-

[3] In this and in the following theorems, the statement can be made more general if x is replaced by $x + \gamma$, where γ is a constant independent of x but having the same unit as x. The effect of this is to place the zero of u at some point different from the zero of x. In psychophysics the constant γ may be regarded as the threshold. The presence of such a constant means, of course, that a plot of log u vs. log x will not in general be a straight line. If, however, the independent variable is measured in terms of deviations from the threshold, the plot may become straight. Such nonlinear plots have been observed, and in at least some instances the degree of curvature seems to be correlated with the magnitude of the threshold. Further empirical work is needed to see whether this is a correct explanation of the curvature.

ent continuum an interval scale, then either $u(x) = \alpha \log x + \beta$, *where* α *is independent of the unit of the independent variable, or* $u(x) = \alpha x^{\beta} + \delta$, *where* β *is independent of the units of both variables and* δ *is independent of the unit of the independent variable.*

Proof. In solving Equation 2, there are two possibilities to consider.

1. If $K(k) \equiv 1$, then define $v = e^{u}$. Equation 2 becomes $v(kx) = D(k)v(x)$, where $D(k) = e^{C(k)} > 0$ and v is continuous, positive, and nonconstant because u is. By Theorem 1, $v(x) = \delta x^{\alpha}$, where α is independent of the unit of x and where $\delta > 0$ because, by definition, $v > 0$. Taking logarithms, $u(x) = \alpha \log x + \beta$, where $\beta = \log \delta$.

2. If $K(k) \not\equiv 1$, then let u and u^* be two different solutions to the problem, and define $w = u^* - u$. It follows immediately from Equation 2 that w must satisfy the functional equation $w(kx) = K(k)w(x)$. Since both u and u^* are continuous, so is w; however, it may be a constant. Since $K(k) \not\equiv 1$, it is clear that the only constant solution is $w = 0$, and this is impossible since u and u^* were chosen to be different. Thus, by Theorem 1, $w(x) = ax^{\beta}$. Substituting this into the functional equation for w, it follows that $K(k) = k^{\beta}$. Then setting $x = 0$ in Equation 2, we obtain $C(k) = u(0) \times (1 - k^{\beta})$. We now observe that $u(x) = \alpha x^{\beta} + \delta$, where $\delta = u(0)$, is a solution to Equation 2:

$$\begin{aligned} u(kx) &= \alpha k^{\beta} x^{\beta} + \delta \\ &= \alpha k^{\beta} x^{\beta} + u(0)k^{\beta} + u(0) - u(0)k^{\beta} \\ &= k^{\beta} u(x) + u(0)(1 - k^{\beta}) \\ &= K(k)u(x) + C(k) \end{aligned}$$

Any other solution is of the same form because

$$\begin{aligned} u^*(x) &= u(x) + w(x) \\ &= \alpha x^{\beta} + \delta + ax^{\beta} \\ &= (\alpha + a)x^{\beta} + \delta \end{aligned}$$

It is easy to see that δ is independent of the unit of x and β is independent of both units.

A much simpler proof of this theorem can be given if we assume that u is differentiable in addition to being continuous. Since the derivative of an interval scale is a ratio scale, it follows immediately that du/dx satisfies Equation 1 and so, by Theorem 1, $\frac{du}{dx} = \alpha x^{\beta}$. Integrating, we get

$$u(x) = \begin{cases} \dfrac{\alpha}{\beta + 1} x^{\beta+1} + \delta & \text{if } \beta \neq -1 \\ \alpha \log x + \delta & \text{if } \beta = -1 \end{cases}$$

Theorem 3. If the independent continuum is a ratio scale and the dependent continuum is a logarithmic interval scale, then either $u(x) = \delta e^{\alpha x^{\beta}}$, *where* α *is independent of the unit of the dependent variable,* β *is independent of the units of both variables and* δ *is independent of the unit of the independent variable, or* $u(x) = \alpha x^{\beta}$, *where* β *is independent of the units of both variables.*

Proof. Take the logarithm of Equation 3 and let $v = \log u$:

$$v(kx) = K^*(k) + C(k)v(x)$$

where $K^*(k) = \log K(k)$. By Theorem 2, either

$$v(x) = \alpha x^{\beta} + \delta^* \text{ or } v(x) = \beta \log x + \alpha^*$$

Taking exponentials, either

$$u(x) = \delta e^{\alpha x^{\beta}} \quad \text{or} \quad u(x) = \alpha x^{\beta}$$

where $\delta = e^{\delta^*}$ and, in the second equation, $\alpha = e^{\alpha^*}$.

Theorem 4. If the independent continuum is an interval scale, then it is impossible for the dependent continuum to be a ratio scale.

Proof. Let $c = 0$ in Equation 4, then by Theorem 1 we know $u(x) = \alpha x^{\beta}$.

Now set $k = 1$ and $c \neq 0$ in Equation 3:

$$\alpha(x + c)^\beta = K(1,c)\alpha x^\beta$$

so

$$x + c = K(1,c)^{1/\beta}x$$

which implies x is a constant, contrary to our assumption that both continua have more than one point.

Theorem 5. If the independent and dependent continua are both interval scales, then $u(x) = \alpha x + \beta$, where β is independent of the unit of the independent variable.

Proof. If we let $c = 0$, then Equation 5 reduces to Equation 2 and so Theorem 2 applies. If $u(x) = \alpha \log x + \beta$, then choosing $k = 1$ and $c \neq 0$ in Equation 5 yields

$$\alpha \log (x + c) + \beta = K(1,c)\alpha \log x + K(1,c)\beta + C(1,c)$$

By taking the derivative with respect to x, it is easy to see that x must be a constant, which is impossible.

Thus, we must conclude that $u(x) = \alpha x^\delta + \beta$. Again, set $k = 1$ and $c \neq 0$,

$$\alpha(x + c)^\delta = K(1,c)\alpha x^\delta + K(1,c)\beta + C(1,c)$$

If $\delta \neq 1$, then differentiate with respect to x:

$$\alpha\delta(x + c)^{\delta-1} = K(1,c)\alpha\delta x^{\delta-1}$$

which implies x is a constant, so we must conclude $\delta = 1$. It is easy to see that $u(x) = \alpha x + \beta$ satisfies Equation 5.

Theorem 6. If the independent continuum is an interval scale and the dependent continuum is a logarithmic interval scale, then $u(x) = \alpha e^{\beta x}$, where α is independent of the unit of the independent variable and β is independent of the unit of the dependent variable.

Proof. Take the logarithm of Equation 6 and let $v = \log u$:

$$v(kx + c) = K^*(k,c) + C(k,c)v(x)$$

where $K^*(k,c) = \log K(k,c)$. By Theorem 5,

$$v(x) = \beta x + \alpha^*$$

so

$$u(x) = \alpha e^{\beta x}$$

where $\alpha = e^{\alpha^*}$.

Theorem 7. If the independent continuum is a logarithmic interval scale, then it is impossible for the dependent continuum to be a ratio scale.

Proof. Let $v(\log x) = u(x)$, i.e., $v(y) = u(e^y)$, then Equation 7 becomes

$$v(\log k + c \log x) = K(k,c)u(\log x)$$

Thus, $\log x$ is an interval scale and v is a ratio scale, which by Theorem 4 is impossible.

Theorem 8. If the independent continuum is a logarithmic interval scale and the dependent continuum is an interval scale, then $u(x) = \alpha \log x + \beta$, where α is independent of the unit of the independent variable.

Proof. Let $v(\log x) = u(x)$, then Equation 8 becomes

$$v(\log k + c \log x) = K(k,c)v(\log x) + C(k,c)$$

so $\log x$ and v are both interval scales. By Theorem 5,

$$u(x) = v(\log x) = \alpha \log x + \beta$$

Theorem 9. If the independent and dependent continua are both logarithmic interval scales, then $u(x) = \alpha x^\beta$, where β is independent of the units of both the independent and dependent variables.

Proof. Take the logarithm of Equation 9 and let $v = \log u$:

$$v(kx^c) = K^*(k,c) + C(k,c)v(x)$$

where $K^*(k,c) = \log K(k,c)$. By Theorem 8,

$$v(x) = \beta \log x + \alpha^*$$

so

$$u(x) = e^{v(x)} = \alpha x^\beta$$

where $\alpha = e^{\alpha^*}$.

ILLUSTRATIONS

It may be useful, prior to discussing these results, to cite a few familiar laws that accord with some of them. The best source of examples is classical physics, where most of the fundamental variables are idealized as continua that form either ratio or interval scales. No attempt will be made to illustrate the results concerning logarithmic interval scales, because no actual use of scales of this type seems to have been made.

The variables entering into Coulomb's law, Ohm's law, and Newton's gravitation law are all ratio scales, and in each case the form of the law is a power function, as called for by Theorem 1. Additional examples of Theorem 1 can be found in geometry since length, area, and volume are ratio scales; thus the dependency of the volume of a sphere upon its radius or of the area of a square on its side are illustrations.

Other important variables such as energy and entropy form interval scales, and we can therefore anticipate that as dependent variables they will illustrate Theorem 2. If a body of constant mass is moving at velocity v, then its energy is of the form $\alpha v^2 + \delta$. If the temperature of a perfect gas is constant, then as a function of pressure p the entropy of the gas is of the form $\alpha \log p + \beta$. No examples, of course, are possible for Theorem 4.

As an example of Theorem 5 we may consider ordinary temperature, which is frequently measured in terms of the length of a column of mercury. Although length as a measure forms a ratio scale, the length of a column of mercury used to measure temperature is an interval scale (subject to the added constraint that the length is positive), since we may choose any initial length to correspond to a given temperature, such as the freezing point of water. If the temperature scale is also an interval scale, as is usually assumed, then the only relation possible according to Theorem 5 is the linear one.

DISCUSSION

Some with whom I have discussed these theorems—which from a mathematical point of view are not new—have had strong misgivings about their interpretation; the feeling is that something of a substantive nature must have been smuggled into the formulation of the problem. They argue that practically any functional relation can hold between two variables and that it is an empirical, not a theoretical, matter to ascertain what the function may be in specific cases. To support this view and to challenge the theorems, they have cited examples from physics, such as the exponential law of radioactive decay or some sinusoidal function of time, which seem to violate the theorems stated above. We must, therefore, examine the ways in which these examples bypass the rather strong conclusions of the present theory.

All physical examples which have been suggested to me as counterexamples to the theorems have a common form: the independent variable is a ratio scale, but it enters into the equation in a dimensionless fashion. For example, some identifiable value of the variable is taken as the reference level x_0, and all other values are expressed in reference to it as x/x_0. The effect of this is to make the quantity x/x_0 independent of the unit used to measure the variable, since $kx/kx_0 = x/x_0$. In periodic functions of time, the period is often used as a reference level. Slightly more generally, the independent variable only appears multiplied by a constant c whose units are the inverse of those of x. Thus, whenever the unit of x is changed by multiplying all values by a constant $k > 0$, it is necessary to adjust the unit of c by multiplying it by $1/k$. But this means that the product is independent of k: $(c/k)(kx) = cx$. The time constant in the law of radioactive decay is of this nature.

There are two ways to view these examples in relation to the principle stated above. If the ratio scale x is taken to be the independent variable, then the invariance part of the principle is not satisfied by these laws. If, however, for the purpose of the law under consideration the dimensionless quantity cx is treated as the variable, then no violation has occurred. Although surprising at first glance, it is easy to see that the principle imposes no limitations upon the form of the law when the independent variable is dimensionless, i.e., when no transformations save the identity are admissible.

We are thus led to the following conclusion. Either the independent variable is a ratio scale that is multiplied by a dimensional constant that makes the product independent of the unit of the scale, in which case there is no restriction upon the laws into which it may enter, or the independent variable is not rendered dimensionless, in which case the laws must be of the form described by the above theorems. Both situations are found in classical physics, and one wonders if there is any fundamental difference between them. I have not seen any discussion of the matter, and I have only the most uncertain impression that there is a difference. In many physical situations where a dimensional constant multiplies the independent variable, the dependent variable is bounded. This is true of both the decay and periodic laws. Usually, the constant is expressed in some natural way in terms of the bounds, as, for example, the period of a periodic function. Whether dimensional constants can legitimately be used in other situations, or whether they can always be eliminated, is not at all apparent to me.

One may legitimately question which of these alternatives is applicable to psychophysics, and the answer is far from clear. The widespread use of, say, the threshold as a reference level seems at first to suggest that psychophysical laws are to be expressed in terms of dimensionless quantities; however, the fact that this is done mainly to present results in decibels may mean no more than that the given ratio scale is being transformed into an interval scale in accordance with Theorem 2:

$$\begin{aligned} y &= \alpha \log x/x_0 \\ &= \alpha \log x + \beta \end{aligned}$$

where

$$\beta = -\alpha \log x_0$$

In addition to dimensionless variables as a means of by-passing the restrictions imposed by scale types, three other possibilities deserve discussion.

First, the idealization that the scales form mathematical continua and that they are related by a continuous function may not reflect the actual state of affairs in the empirical world. It is certainly true that, in detail, physical continua are not mathematical continua, and there is ample reason to suspect that the same holds for psychological variables. But the assumptions that stimuli and responses both form continua are idealizations that are difficult to give up; to do so would mean casting out much of psychophysical theory. Alternatively, we could drop the demand that the function relating them be continuous, but it is doubtful if this would be of much help by itself. The discontinuous solutions to, say, Equation 1 are manifold and extremely wild in their behavior. They are so wild that it is difficult to say anything precise about them at all (see Hamel, 1905; Jones: 1942a, 1942b), and it is doubtful that such solutions represent empirical laws.

Second, casual observation suggests that it might be appropriate to assume that at least the dependent variable is bounded, e.g., that there is a psychologically maximum loudness. Although plausible, boundedness cannot be imposed by itself since, as is shown in the theorems, all the continuous solutions to the appropriate functional equations are unbounded if the functions are increasing, as they must be for empirical reasons. It seems clear that boundedness of the dependent variable is intimately tied up either with introducing a reference level so that the independent variable is an absolute scale or with some discontinuity in the formulation of the problem, possibly in the nature of the variables or possibly in the function relating them. Actually, one can establish that it must be in the nature of the variables. Suppose, on the contrary, that the variables are ratio scales that form numerical continua and that they are related by a function u that is nonnegative, nonconstant, and monotonic increasing, but not necessarily continuous. We now need only show that u cannot be bounded to show that the discontinuity must exist in the variable. Suppose, therefore, that it is bounded and that the bound is M. By Equation 1, $u(kx) = K(k)u(x) \leq M$, so $u(x) \leq M/K(k)$. For $k \geq 1$, the monotonicity of u implies that $u(x) \leq u(kx) = K(k)u(x)$, so choosing $u(x) > 0$ we see that $K(k) \geq 1$. If for some $k \geq 1$, $K(k) > 1$, then K can be made arbitrarily large since, for any integer n, $K(k^n) = K(k)^n$, but since $u(x) \leq \frac{M}{K(k)}$, this implies $u \equiv 0$, contrary to assumption. Thus, for all $k \geq 1$, $K(k) = 1$, which by Equation 1, means $u(kx) = u(x)$, for all x and $k \geq 1$. This in turn implies u is a constant, which again is contrary to assumption. Thus, we have established our claim that some discontinuity must reside in the nature of the variables.

Third, in many situations, there are two or more independent variables; for example, both intensity and frequency determine loudness. Usually we hold all but one variable constant in our empirical investigations, but the fact remains that the others are there and that their presence may make some difference in the total range of possible laws. For example, suppose there are two independent variables, x and y, both of which form ratio scales and that the dependent variable u is also a ratio scale, then the analogue of Equation 1 is

$$u(kx,hy) = K(k,h)u(x,y)$$

where $k > 0$, $h > 0$, and $K(k,h) > 0$. We know by Theorem 1 that if we hold one variable, say y, fixed at some value and let $h = 1$, then the solution must be of the form

$$u(x,y) = \alpha(y)x^{\beta(y)}$$

But holding x constant and letting $k = 1$, we also know that it must be of the form

$$u(x,y) = \delta(x)y^{\epsilon(x)}$$

Thus,

$$\alpha(y)x^{\beta(y)} = \delta(x)y^{\epsilon(x)}$$

If we restrict ourselves to u's having partial derivatives of both variables, this equation can be shown (see Section 2.C.2 of Luce [in press]) to have solutions only of the form:

$$u(x,y) = ax^b y^{c+d \log x}$$

Thus, the principle again severely restricts the possible laws, even when we admit more than one independent variable.[4]

It must be emphasized that the remark in Footnote 3 does not apply here. If a function that depends upon one independent variable is added to the other, e.g.,

$$u(x,y) = \alpha(y)[x + \gamma(y)]^{\beta(y)}$$

then wholly new solution possibilities exist (see Section 2.C.3 of Luce [in press]).

In sum, there appear to be two ways around the restrictions set forth in the theorems. The first can be viewed either as a rejection of Part 2 of the principle or as the creation of a dimensionless independent variable from a ratio scale; it involves the presence of dimensional constants that cancel out the dimensions of the independent variables. This appears to be particularly appropriate if the dependent variable has a true, well-defined bound. The second is to reject the idealization of the variables as numerical continua and, possibly, to assume that they are bounded.

On the other hand, if the theorems are applicable, then the possible psychophysical (and other) laws become severely limited. Indeed, they are so limited that one can argue that the important question is not to determine the forms of the laws, but rather to create empirically testable measurement theories for the several psychophysical methods in order that we may know for certain what types of scales are being obtained. Once this is known, the form of the psychophysical functions is determined except for some numerical constants. In the meantime, however, experimental determinations of the form of the psychophysical functions by methods for which no measurement theories exist provides at least indirect evidence of the type of scale being obtained. For example, the magnitude methods seem to result in power functions, which suggests that the psychological measure is either a ratio or logarithmic interval scale, not an interval scale. Since the results from cross-modality matchings tend to eliminate the logarithmic interval scale as a possibility, there is presumptive evidence that these methods yield ratio scales, as Stevens has claimed.

SUMMARY

The following problem was considered. What are the possible forms of a substantive theory that relates a dependent variable in a continuous manner to an independent variable? Each variable is idealized as a nu-

[4] The use of this argument to arrive at the form of $u(x,y)$ seems much more satisfactory and convincing than the heuristic development given in Section 2.C of Luce (in press), and the empirical suggestions given there should gain correspondingly in interest as a result of the present work.

TABLE 2

THE POSSIBLE LAWS SATISFYING THE PRINCIPLE OF THEORY CONSTRUCTION

Scale Types		Possible Laws	Comments[a]
Independent Variable	Dependent Variable		
ratio	ratio	$u(x)=\alpha x^{\beta}$	β/x; β/u
ratio	interval	$u(x)=\alpha \log x+\beta$	α/x
		$u(x)=\alpha x^{\beta}+\delta$	β/x; β/u; δ/x
ratio	log interval	$u(x)=\delta e^{\alpha x^{\beta}}$	α/u; β/x; β/u; δ/x
		$u(x)=\alpha x^{\beta}$	β/x; β/u
interval	ratio	impossible	
interval	interval	$u(x)=\alpha x+\beta$	β/x
interval	log interval	$u(x)=\alpha e^{\beta x}$	α/x; β/u
log interval	ratio	impossible	
log interval	interval	$u(x)=\alpha \log x+\beta$	α/x
log interval	log interval	$u(x)=\alpha x^{\beta}$	β/x; β/u

[a] The notation α/x means "α is independent of the unit of x."

merical continuum and is restricted by its measurement theory to being either a ratio, an interval, or a logarithmic interval scale. As a principle of theory construction, it is suggested that transformations of the independent variable that are admissible under its measurement theory shall not result in *inadmissible transformations* of the dependent variable (consistency) and that the form of the functional relation between the two variables shall not be altered by admissible transformation of the independent variable (invariance). This principle limits significantly the possible laws relating the two continua, as shown in Table 2.

These results do not hold in two important circumstances. First, if the independent variable is a ratio scale that is rendered dimensionless by multiplying it by a constant having units reciprocal to those of the independent variable, then either the principle has no content or it is violated, depending upon how one wishes to look at the matter. Second, if the variables are discrete rather than continuous, or if the functional relation is discontinuous, then laws other than those given in Table 2 are possible.

REFERENCES

BUSH, R. R., MOSTELLER, F., & THOMPSON, G. L. A formal structure for multiple-choice situations. In R. M. Thrall, C. H. Coombs, & R. L. Davis (Eds.), *Decision processes.* New York: Wiley, 1954. Pp. 99–126.

COHEN, M. R., & NAGEL, E. *An introduction to logic and scientific method.* New York: Harcourt, Brace, 1934.

HAMEL, G. Eine Basis aller Zahlen und die unstetigen Lösungen der Funktionalgleichung: $f(x+y)=f(x)+f(y)$. *Math. Annalen,* 1905, **60**, 459–462.

JONES, F. B. Connected and disconnected plane sets and the functional equation $f(x)+f(y)=f(x+y)$. *Bull. Amer. Math. Soc.,* 1942, **48**, 115–120. (a)

JONES, F. B. Measure and other properties of a Hamel basis. *Bull. Amer. Math. Soc.,* 1942, **48**, 472–481. (b)

KEMENY, J. G., & THOMPSON, G. L. The effect of psychological attitudes on the outcomes of games. In M. Dresher, A. W. Tucker, & P. Wolfe (Eds.), *Contributions to the theory of games, III.* Princeton: Princeton Univer. Press, 1957. Pp. 273–298.

LUCE, R. D. *Individual choice behavior: A theoretical analysis.* New York: Wiley, in press.

LUCE, R. D., & EDWARDS, W. The derivation of subjective scales from just noticeable differences. *Psychol. Rev.,* 1958, **65**, 222–237.

MOSTELLER, F. The mystery of the missing corpus. *Psychometrika,* 1958, **23**, 279–289.

STEVENS, S. S. On the theory of scales of measurement. *Science,* 1946, **103,** 677–680.
STEVENS, S. S. Mathematics, measurement and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 1–49.
STEVENS, S. S. On the averaging of data. *Science,* 1955, **121,** 113–116.
STEVENS, S. S. The direct estimation of sensory magnitudes—loudness. *Amer. J. Psychol.,* 1956, **69,** 1–25.
STEVENS, S. S. On the psychophysical law. *Psychol. Rev.,* 1957, **64,** 153–181.
STEVENS, S. S. Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *J. exp. Psychol.,* 1959, **57,** 201–209.
STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.,* 1957, **54,** 377–411.
THURSTONE, L. L. A law of comparative judgment. *Psychol. Rev.,* 1927, **34,** 273–286.
VON NEUMANN, J., & MORGENSTERN, O. *The theory of games and economic behavior.* (2nd ed.) Princeton: Princeton Univer. Press, 1947.

(Received December 2, 1958)