

*On the Status of "Direct"
Psychophysical Measurement*

Whether the water in a pot feels cold, cool, warm, or hot depends both on how long the pot has been on the fire and how long the testing hand has been in from the cold. In the recognition of such primitive facts one can discern the beginnings of both physics and psychophysics.

The fundamental invariances of physics began to emerge in pure, mathematically expressible form only when it became possible (a) to refine the scale of descriptive labels that could be applied to an object under study (e.g., from four such crude labels as "cold," "cool," "warm," and "hot" to a hundred or more distinguishable

Note: I first set forth the basic ideas underlying this essay over a decade ago in an unpublished but rather widely circulated note dated November 15, 1966 and titled "What does the psychophysicist measure?" Principally, these ideas were (a) that psychophysical judgments are essentially relative judgments, (b) that magnitude estimation and cross-modality matching determine no more than the ordinal structure of internal magnitudes, and (c) that what the psychophysicist measures on a "ratio scale" is a parameter that characterizes each sensory continuum — not the magnitude of any one sensation within such a continuum. Except for minor editing, the addition of some more recent references, and a number of deletions, the present paper is essentially identical to the draft of a more extensive development of these ideas that I prepared, under the present title, before leaving Harvard in June of 1968. It represents my most ambitious attempt to come to terms with the psychophysical claims of my late Harvard colleague, Professor S. S. Stevens. Regrettably, following my shift in geographical location and field of research, I had not until now found the occasion to return to the task of revising this manuscript for publication. In the meantime other writers have further explored some of these ideas. I believe the most elegant (and generous) published statements along these lines to be those of David H. Krantz (1972a, b). As I note in the relevant section, some of the formalizations presented in the present paper owe much to suggestions made to me by David V. Cross (personal communication, 1968). I want also to acknowledge the support provided by the National Science Foundation (Grants GS-1302 and BMS 75-02806) during the preparation and revision of this essay.

levels in the height of a column of mercury), and (b) to ensure that the descriptive label obtained from such a refined scale would relate more to the state of the object itself and less to the extraneous state of the measurer or his measuring instrument.

Once physics had attained this degree of quantitative precision and independence of the imperfectly correlated reactions of the human observer, the question arose as to whether these purely “psychological” responses—now that they had been conceptually distinguished from their purely physical correlates—might not be found to possess a certain kind of order of their own. If so, one could begin to contemplate a science of *psychophysics*; that is, a science that would have as its goal the quantitative specification of any relations that hold between such a psychological order, on the one hand, and the independently established physical order, on the other. I have attempted in this chapter to clarify the extent to which we have approached this goal.

The Problem of the Construction of Psychophysical Scales

In the development of purely physical science, mathematically formulated laws of predictive precision and generality were of course made possible only through the extensive refinement of physical scales of measurement such, for example, as the scale of temperature. It was therefore natural, in attempting the subsequent development of a psychophysical science, to strive for a similar refinement of psychological scales of measurement. For, as long as the responses of the human observer were limited to just a few ill-defined qualitative labels, the relations between these responses and the physical variables could not approach the kind of precision and generality characteristic of the laws relating purely physical variables.

CONSTRUCTION OF SCALES FOR THE MEASUREMENT OF PHYSICAL MAGNITUDES

In the case of physics, the perfection of scales of measurement was critically dependent upon the development of theory. The earliest temperature-sensitive devices, such as the “thermoscope” or “weather glass” of 1600, provided only a somewhat more purified and objective way of gauging “warmth.” The still largely intuitive notion of temperature had not yet acquired sufficient theoretical

articulation to support more than an ordinal structure. That is, although it could consistently be determined whether one object was warmer or cooler than another (even when the difference was quite small), there was little basis for asserting anything further about the quantitative magnitude of such a difference in temperature. Indeed, even when the thermoscope was transformed into the first crude thermometer by the addition of a graduated scale some time around 1610 (Middleton, 1966), the scale was necessarily quite arbitrary and so did not really provide an adequate basis for assigning numerical values to temperatures in any unique way.

It was the considerably later developments of thermodynamics and the kinetic theory of heat that furnished the structure necessary for the transformation of the original, intuitive notion of warmth into the present, fully articulated concept of temperature. It was only by recourse to abstract thermodynamic arguments concerning "ideal heat engines" (see, e.g., Becker, 1967, pp. 18-20) that in 1848 William Thomson (Lord Kelvin) was able to present a satisfactory rationale for the specification of temperatures by numerical values that were uniquely determined except for multiplication by a positive constant (the arbitrary unit of measurement that determines whether we are talking, say, about degrees Kelvin or degrees Rankine).

On the resulting "thermodynamic scale" the difference between two temperatures can be defined in terms of the mechanical work extractable by a Carnot engine working between the two temperatures. Thus it became for the first time fully meaningful to say that the difference in temperature between two cool objects, *A* and *B*, is equal to the difference in temperature between two warm objects, *C* and *D*. In terms of the classificatory scheme of scale types set forth by Stevens (1946, 1951), the merely "ordinal scale" of warmth had at this point become an "interval scale" of temperature. Further, with the conceptualization of the absolute zero point of temperature, which can only be approached but never attained (and which, according to the kinetic theory, corresponds to the cessation of all relative molecular motion), it became meaningful to talk about ratios as well as differences. Of two objects, *A* and *B*, not only could one say that the temperature of *B* is greater than the temperature of *A*, one could further state that it is twice as great, three times as great,

or whatever the case might be. At this point, the scale of temperature finally emerged as what would be classified in Stevens's scheme as a full-fledged "ratio scale."

Apparently, then, the interval and ratio properties of the thermodynamic scale of temperature can be fully justified only by reference to physical theory. It is of course a convenient (and not entirely adventitious) outcome that the equally spaced graduations on the laboratory thermometer correspond to nearly (though, significantly, not precisely) equal differences on the thermodynamic scale. But even the hydrogen thermometer, which most closely approximates the theoretically ideal thermodynamic scale, is subject to correction owing to the departure of hydrogen from a perfect gas (with the consequence that the desired "thermometric" property is approximated only as the product of pressure times volume is extrapolated to zero pressure). Again, the real reason for specifying temperature on this scale (and for imposing slight corrections on the equally spaced graduation of the thermometer) is the resulting simplification in the mathematical structure of physical theory as a whole. For this theory is concerned not just with temperature, but also with a host of other variables and concepts that are related to temperature and to each other in a vast, interdigitated complex.

In the last analysis, of course, the justification for the whole theoretical edifice must be sought in the account that it provides for concrete empirical observations. Nevertheless, it appears that the variables (such as temperature) that are measured by means of physical scales as now refined are themselves essentially of the nature of theoretical constructs. For, to the extent that we identify the temperature of a macroscopic body with the kinetic energy of its microscopic constituents, the temperature itself recedes from the realm of concrete, directly experienceable entities.

CONSTRUCTION OF SCALES FOR THE MEASUREMENT OF PSYCHOLOGICAL MAGNITUDES

I have argued that refined scales for the measurement of purely physical quantities originally evolved out of attempts to eliminate the influence of the variable internal states of the observer upon his evaluations of the state of an external object. This suggests the possibility of using a reverse strategy to develop comparably refined

scales for the measurement of what might be called purely psychological magnitudes; namely, those very states of the observing subject that, in the process of perfecting physical measurement, had finally been cast entirely aside.¹

Within the original use of such crude, descriptive labels as "hot" and "cold," that is, one can discern the precursors of two quite distinct concepts: (a) that of a purely physical magnitude (temperature) that is conceived as residing entirely within the external object, and (b) that of a purely psychological magnitude (perceived warmth) that is conceived as residing entirely within the observing subject. Indeed, either magnitude can exist quite independently of the other. The physical temperatures of objects can be registered automatically without the occurrence of any sensation of warmth, and (during direct electrical excitation of the cortex, spontaneous hallucinations, or dreams) a sensation of warmth can occur in the total absence of a corresponding physical stimulus.

It might seem that I have ignored a fundamental asymmetry between the two types of magnitudes. For even if physical temperature is a theoretical construct, inaccessible to direct observation, the corresponding sensation of warmth is surely directly experienced by the subject who reports its occurrence. However, with the adoption of the behavioristic orientation toward sensations required for the intersubjective development of a public science, statements about sensations assume fundamentally the same status as statements about the theoretical variables of physics. If the concept of a sensation of warmth is introduced into behavioral theory, it is for the same reason that the concept of a temperature is introduced into physical theory—not because it is demanded by any single observation, but because it simplifies the theory as a whole. Just as a temperature is identified, theoretically, with a certain unobserved state of molecular agitation in the external object, a sensation of warmth would presumably be identified, theoretically, with some unobserved pattern or level of neuronal activity in the brain.

Now physical measurement owes its great power not only to its independence of the variable state of the observer but also to its high degree of quantitative precision. To what extent can this same kind of precision be achieved in the measurement of purely psychological variables—variables such as visual brightness, auditory loudness, or

tactual warmth? Such sensations do seem to have certain definite subjective magnitudes. But to what extent can we specify explicit procedures of measurement that will enable us to pin such magnitudes down on anything approaching the refined structure of a numerical scale?

The history of physical measurement indicates that, if this is to be accomplished, its accomplishment will hinge critically upon the available theoretical structure. A subject can of course say more about his thermal sensations than merely whether each is "cold," "cool," "warm," or "hot." He can, for example, report the judgment that one such sensation seems about twice as strong (i.e., hot) as another. But we cannot safely conclude that the one sensation is, in fact, twice the other just because the subject makes such a verbal report—any more than we could conclude that the temperature is twice the other just because the mercury rises to twice the height. The direct, quantitative indication of either the subject or the thermometer might even be true, but, in the one case as much as in the other, a more than trivial theory is needed to substantiate the claim.

ROLE OF PSYCHOPHYSICAL RELATIONS IN THE CONSTRUCTION OF SCALES

A physical magnitude such as temperature evidently derives much of its significance from its relations with other physical variables. Similarly, if procedures could be devised for making quantitative determinations of psychological magnitudes, the most significant advance in our understanding would come not from the mere measurement of this or that sensation per se. Rather, it would come from the relations that might thus be found to hold between such psychological magnitudes and other measurable variables. The enterprise becomes *psychophysical* (rather than purely psychological) to the extent that some of the most important of these relations are with physical variables.

So far, psychophysical investigations have endeavored principally to determine the functional form of one particular type of relation; namely, the relation between the psychological strength of a sensation (such as perceived loudness, brightness, or warmth) and the physical intensity of the eliciting proximal stimulus (measured, say,

in terms of energy—whether acoustic, electromagnetic, or kinetic). Before such a determination can be made, however, it would seem necessary to meet two conditions.

First, other physical variables that might also influence the psychological magnitude under study must be carefully controlled. One such variable (emphasized at the outset) is the intensity of stimulation to which the receptive organ has previously been adapted. However, although such control may not always be easy in practice, it does not appear to depend upon the resolution of any major theoretical issues and so will not be further considered here.

The second condition is quite another matter. It is that the psychological magnitude in question must be measurable on a scale with more than a merely ordinal structure. We must be able to find out more about a sensation than merely whether it exceeds or falls short of some other sensation. Otherwise, the findings could as well be accounted for by *any* monotonic function of physical intensity. And yet, just how the necessary further structure can be secured remains an issue of continuing dispute.

Actually, as the example of the temperature scale suggests, it is simplistic to suppose that one first erects a fully structured scale and *then* proceeds to relate the variable measured on this scale to other variables. Rather, what seems to have happened in physics is a bi-directional, mutually constraining interaction between these two processes. It was, in fact, the attempt to perfect and simplify the relations with other variables that prescribed what form the internal structure of the temperature scale itself would finally have to take. Perhaps, then, the development of psychophysics can progress only by means of a similar kind of cooperative alternation between the tightening of the structure of a scale and the simplifying of its relations with other variables.

DISCRIMINABILITY AS A BASIS FOR PSYCHOPHYSICAL SCALING

In view of the difficulties that have appeared to confront any attempt to measure psychological magnitudes directly, some psychophysicists, notably Fechner and Thurstone, have turned to discriminability as a method for inferring psychological magnitudes, so to speak, indirectly. In practice, all such approaches evidently depend upon the variability of subjects' responses and, indeed, de-

pend upon some specific assumption about the relation between that variability and psychological magnitude.

Although he did not conceive of it in precisely these terms, Fechner (1860), who is generally credited with the founding of psychophysics, assumed in effect that this variability is constant and independent of magnitude on the underlying psychological scale. Thus he was led to identify the difference in the psychological magnitudes of two intensities on the same sensory continuum with what amounts to the number of intervening differences in intensity that could be discriminated some specified fraction of the time (i.e., the number of intervening “just noticeable differences” or “*jnds*”).

In order to secure a ratio scale, then, Fechner had only to assume that a zero (or subthreshold) physical intensity leads to a zero psychological magnitude. For, on the resulting scale, one can numerically specify both the *difference* between two sensations (the number of *jnds* from the one to the other) and the *ratio* between those sensations (the number of *jnds* from zero to the one divided by the number of *jnds* from zero to the other).

In his much later approach to psychophysics, Thurstone (1927) was more explicit about its dependence upon the variability of the subject’s responses. In particular, he expressly postulated that the same external stimulus leads to an internal psychological magnitude, the “discriminal process,” that varies from occasion to occasion. Moreover, in a manner that strongly foreshadowed the modern theory of signal detection (Green & Swets, 1966), he went on to specify a decision rule, according to which the subject’s overt response to a particular stimulus was determined by whether the psychological magnitude to which it gave rise on a given occasion exceeded or fell short of some internal criterion magnitude.

However, although he thus provided an explicit mechanism to account for Fechner’s *jnd*, Thurstone himself preferred to take a somewhat different aspect of the subject’s variability as fundamental for the construction of psychological scales. Specifically, instead of assuming that the distribution of the discriminable process has a fixed variance on the underlying psychological scale, he assumed that this distribution has a fixed functional form — namely, the form of the normal (or Gaussian) error function. Under this assumption, the normal distributions of discriminational processes arising from dif-

ferent stimuli could have different variances; in that case, equally often noticed differences would not necessarily correspond to equal differences in psychological magnitude.

Scaling methods in considerable variety have evolved out of Fechner's and Thurstone's original attempts to base psychophysics on discriminability. Essentially these methods are designed to find a spacing of stimuli on the "psychological scale" such that the overt responses of subjects can best be accounted for by some model of the Thurstonian type in which the hypothetical distributions of psychological magnitudes all have either (a) the same variance, as suggested by Fechner, (b) the normal functional form, as proposed by Thurstone, or perhaps (c) merely the same, unspecified functional form (Klemmer & Shrimpton, 1963; Kruskal & Shepard, 1974, p. 154; Levine, 1970, 1972; Shepard, 1965), or even (d) just an optimum degree of smoothness or "continuity" (Carroll, 1963; Shepard & Carroll, 1966).

Since these methods lead to interval or ratio scales (i.e., to scales that are determined essentially to within a linear or even a similarity transformation), they possess enough structure to support some determination of the "psychophysical function." Fechner, as I noted, claimed to have what amounts to a ratio scale of psychological magnitude. All he then needed was the empirical fact, already established by Weber, that the size of a *jnd* in units of physical intensity is approximately proportional to intensity. From this empirical invariance (since known as "Weber's Law"), Fechner deduced that the psychological magnitude of a sensation, as he defined it, must be a logarithmic function of physical intensity.

Fechner's proposed solution to the psychophysical problem appears rather weak, however, when we compare his scale of psychological magnitude with physical scales, such as that for temperature. For the structure of the thermodynamic scale of temperature was, so to speak, conferred on that temperature scale—not just by one relation, but by a richly interconnected web of mutually reinforcing theoretical and empirical relations. By contrast, the structure of Fechner's scale was dictated by the attempt to secure just one relation; viz., the relation that a just noticeable difference subtend the same separation in psychological magnitude regardless of its position on the scale.

This weakness is not, however, inherent in the general choice of discriminability as a basis for psychophysics. Within the broader, Thurstonian framework there is at least the possibility of establishing mutually reinforcing relations among diverse kinds of experimental tasks. It could for example happen that one scale of psychological magnitude would be found to provide a simple and unified account of (a) the distribution of stimuli classified as "same" or "different" or as "greater than" or "less than" any given stimulus (Woodworth, 1938; Torgerson, 1958), (b) the pattern of errors made in tasks of absolute identification or paired-associate learning (Garner, 1952; Luce & Galanter, 1963; Shepard, 1958b), (c) the shape of the "gradient" of stimulus generalization (Guttman & Kalish, 1956; Shepard, 1965) or of discriminative reaction time (Curtis, Paulos, & Rule, 1973; Falmagne, 1971; Shepard, Kilpatrick, & Cunningham, 1975, pp. 127 ff.; Welford, 1960), (d) the way in which accuracy of comparative judgment decays with the delay of the second stimulus (Shepard, 1958a; Wickelgren, 1969), and even (e) phenomena of stimulus-response compatibility (Shepard, 1961) and of classification learning (Shepard & Chang, 1963; Shepard, et al., 1975, pp. 134-135). Then, surely, one could begin to have some confidence in such a psychological scale. (See in particular the more recent work of Falmagne, 1971.)

The extent to which such a synthesis can eventually be achieved for these various sorts of discrimination data appears to be a largely empirical question. In any case the logical status of such a discriminability scale now seems reasonably clear. So the remainder of this paper will be devoted to the examination of the entirely different proposal that, quite apart from the possibility of inferring psychological magnitudes indirectly on the basis of discriminability, we can — and perhaps should — obtain the desired estimates more directly by simply asking the subject himself for a straightforward, quantitative judgment of these magnitudes.

DIRECT JUDGMENT AS A BASIS FOR PSYCHOPHYSICAL SCALING

The late S. S. Stevens, who was undoubtedly the foremost proponent of psychophysics in recent years, launched a sustained attack against the whole attempt to base psychophysics on discriminability. He pointed out that, in physics, the analogue of discriminability

would be error variability—the converse of resolving power. But (except when we approach the irreducible limitations of quantum physics) resolving power is clearly more dependent upon the particular measuring instrument used than it is upon the underlying physical quantity to be measured. Thus the recommendation that we adjust our scale of psychological magnitude solely in order that error variability will be constant across the scale would be like calibrating a scale of electric current just so that the inaccuracy in the deflection of the pointer will be constant across the scale of the galvanometer. But in this latter case (as in the case of temperature), it is relations other than those having to do with such variability that furnish the most fundamental basis for an operational definition of the underlying physical quantity. Possibly this is true as much for the definition of the strength of a sensation as for the definition of the strength of an electric current. Certainly it can be seen as a curious aspect of scaling procedures based on discriminability that they essentially depend upon the presence of what we usually seek to minimize—namely, variability or error.

At a turning point in his long endeavor to develop a method for estimating psychological magnitudes without depending on variability, Stevens (1956) proposed to have subjects simply report direct, quantitative estimates of the strengths of their sensations. In the instructions to the subject, he explicitly specified that the numbers given should be proportional to the strengths of the corresponding sensations. Suppose, for example, that the first stimulus had been called "40." Then the instructions evidently required that, if the psychological magnitude of the second stimulus seemed twice as great as that of the first, it should be called "80"; if half as great, "20"; and so on.

Now the responses obtained from subjects in this method of "direct magnitude estimation" have manifested an undeniable degree of consistency and order for a wide class of sensory continua; namely, the class of what Stevens (1957) has called "prothetic" continua. Principally this class includes the "intensive" and "extensive" dimensions of stimuli along which variation is perceived as the purely *quantitative* addition or subtraction of the same homogeneous quality, as in loudness, brightness, or length. This class is thus contrasted with the class of "metathetic" continua along which variation

is perceived, rather, as the *qualitative* substitution of one kind of experience for another, as in auditory pitch or, perhaps, visual hue.

If we confine our attention specifically to “prothetic” continua, then, we find that the (geometric) mean of the numbers produced by the subjects for each stimulus closely approximates a power function of the physical intensity of the stimulus. Moreover, although the numerical value of the exponent in the best-fitting power function does vary somewhat from subject to subject, when the values are averaged over even a small group of subjects, the resulting average value turns out to be a rather stable characteristic of the particular sensory continuum. Such average values range from between 0.3 to 0.6, for the continuum of intensity of a light (depending on the visual size of the source), all the way up to 4.5, for the continuum of intensity of an electric shock (Stevens, 1966a).

On the basis of these results, Stevens (1961) proclaimed that the logarithmic function that Fechner had derived for the central psychophysical relation should be rejected in favor of a power function (a conclusion, incidently, that he notes had been first advanced and then withdrawn, without the benefit of sufficient evidence, some 100 years earlier by Fechner’s contemporary, Plateau).

Even with the large body of evidence that Stevens and his followers have amassed, however, the claims that the power function rather than the logarithmic function is in fact the true psychophysical law and that one can, and should, now measure sensations by direct magnitude estimation have continued to come under assault (e.g., Graham, 1958; Helson, 1964; Luce, 1972; Savage, 1970; Torgerson, 1960; Treisman, 1963, 1964). For the most part, these attacks seem to be aimed at the implications that Stevens claims to draw from this evidence, more than at the empirical evidence itself.

With respect to the empirical evidence, it is generally conceded that systematic departures from the overall power function are found near threshold, near other “anchoring” background or reference stimuli, or when the stimuli to be judged are themselves drawn from a set of intensities that is unevenly spaced, bunched, or truncated on the “true” underlying psychological scale. But such departures seem susceptible to correction in principle—for example, by the rational introduction of an appropriate, empirically estimable threshold constant, by the avoidance of undesired “anchors,” and by “iter-

ative" experiments designed to converge on an optimum distribution or spacing of stimulus intensities (e.g., see Pollack, 1965). What follows here, in any case, will take the validity of the basic empirical generalizations for granted and will focus instead on the implications of these generalizations for the measurement of psychological magnitudes and for the development of psychophysical theory.

SOME APPARENT LIMITATIONS OF DIRECT MAGNITUDE ESTIMATION

The first problem that we confront here is basically like the problem that the mere affixing of a numbered scale to a thermoscope provided no sound basis for saying whether one temperature was twice or three times another. Indeed it later became apparent that the numbers inscribed on these early scales very definitely were *not* proportional to the corresponding absolute temperatures required by physical theory. Without a theory, then, how can we assume that the numbers proffered by a subject—any more than the numbers indicated on the arbitrary scale of the thermoscope—are proportional to any underlying quantity?

We cannot answer this question merely by insisting that the subject is expressly instructed to give a number that *is* proportional to the underlying psychological magnitude. For, in the absence of any independent access to that psychological magnitude, how could we be certain that the subject is following our instruction? How, indeed, could we ever have taught the subject to make such reports correctly in the first place? Surely it would be a risky business to assume, just because an instruction was issued, that it was followed. (What if we were to instruct the subject to repeat back a 24-digit number, or to report the direction of a weak magnetic field?)

The difficulty inherent in this "direct" approach can be brought out in even clearer, mathematical terms as follows: According to Stevens (1957) the fundamental psychophysical relation is of the general form

$$\Psi = f_1(S), \quad (1)$$

where S is the physically measured intensity of a stimulus, Ψ is the resulting psychological magnitude, and f_1 is the "psychophysical" function that transforms one into the other. Clearly, though, this formulation is incomplete. It takes us only from the external, measurable stimulus, S , into the internal, unobservable sensation, Ψ .

In order to complete the formulation and to give it any empirical content, we must get back out to the externally recordable response, R , by means of a second, reverse “psychophysical” transformation

$$R = f_2(\Psi). \quad (2)$$

But, as has previously been noted (e.g., by Treisman, 1964), since the intervening variable, Ψ , is not itself observable, the responses of the subject can at most determine the form of the single, overall “physical-physical” relation

$$R = f_3(S) = f_2\{f_1(S)\}, \quad (3)$$

in which Ψ does not explicitly appear.

Apparently, unless we arbitrarily fix one of the two component functions of f_3 (viz., either f_1 or f_2), both of the two component functions must remain wholly unknown. Hence the conclusion that Stevens has drawn from the results of magnitude estimation—namely, that the function f_1 is a power function—evidently depends upon the implicit assumption that the instructions have sufficed to ensure that f_2 is itself a power function (hopefully with an exponent of unity). But the grounds for assuming that the instructions would have precisely this effect seem never to have been adequately explained.

In fact the situation is even worse than this. For what the subject really gives us in a magnitude estimation is, after all, only a discrete, learned response (i.e., a word). It is not anything that itself even possesses a definite quantitative magnitude. So, as a number of psychophysicists have noted, whether it is legitimate to speak of “ratios” or “differences” between these verbal responses or, indeed, to compute their geometric means is not something that can safely be taken for granted (Garner, 1954; Graham, 1958, p. 68; Luce & Galanter, 1963, p. 274; McGill, 1960, p. 67; Oyamo, 1968); it needs to be justified.²

But those who have considered how a child might learn to use such phrases as “twice as great,” “three times as great,” etc., have tended to argue that the relations among the public or objective quantities, S , rather than the relations among the private or subjective magnitudes, Ψ , generally furnish the criteria for correct use (Skinner, 1945, 1957; Treisman, 1964; Warren, 1958; Wittgenstein,

1953). Thus, in a situation of comparing sticks of different lengths, the child may learn that the phrase “twice as long” is appropriate when two of the equal, shorter sticks laid end-to-end reach just as far as the single, longer stick.

It may even be that what he really learns is to infer these objective relations among the physical quantities S from the subjective relations among his immediately given psychological magnitudes Ψ . But no matter what these latter, purely subjective relations may be, his commerce with the public, physical world is likely, in the words of Thouless (1931), to favor a “regression,” in his overt responses, to the “real objects.” Whatever the relevant psychological distances, someone who is attempting to jump from stone to stone across a stream had better regulate each response so that it will carry him over the appropriate *physical* distance. It is not perhaps too surprising, then, that the empirically determined exponent in the power law for visual extent is found to be close to unity (Stevens & Guirao, 1963; Teghtsoonian, 1965).

Just how subjects learn to make judgments on intensive, as opposed to extensive, continua is not entirely clear. It is however a suggestive fact (Stevens, 1960, p. 64) that exponents close to $\frac{1}{2}$ rather than close to 1 have been obtained for intensities presented to the “distance” receptors of vision (0.5 for the brightness of a point source), audition (0.6 for the loudness of a binaural tone), and olfaction (0.55 for the strength of the odor of coffee). Possibly the judgment of an intensive magnitude is in part based on the appreciation of an extensive magnitude. Certainly, if a source that is twice as far away is called half as bright, loud, or strong, then the inverse square law might help to explain the clustering of these fitted exponents about the value 0.5 (Treisman, 1964; Warren, 1958; Warren, Sersen, & Pores, 1958).

On the other hand, one must agree with Stevens (1964) that it seems unlikely that each individual subject has had the opportunity to learn to apply numbers in this way to each separate continuum, *de novo*—particularly in the case of the more novel continua that have been studied, such as strength of electric shock or apparent viscosity (Stevens & Guirao, 1964). Possibly, then, much of this “learning” has been accomplished, not in the individual subject, but in the preceding biological evolution of higher organisms in general

(cf. Shepard, 1964, pp. 63-65; 1975, pp. 96, 115). In this connection there is reason to suppose that output transformations are more readily learnable if the function f_2 in Equation (2) is of a suitably standard or "simple" form (Carroll, 1963). If so, natural selection would tend to favor the development of input transformations in which the resulting subjective magnitude represents an appropriate "phenomenal regression," if not always to a "real object," at least to the physical magnitude relevant to the widest range of adaptive responses.

Regardless of how subjects come to be able to assign numbers consistently to stimuli differing along a sensory continuum, the fact remains that what this assignment directly reveals is not the form of the transformation of S into Ψ , but rather the form of the complete transformation of S into R . Here we come back, then, to the essential dependence of measurement on theory. For surely we cannot safely turn about and take the numerical face value of the overt response R as a direct measure of the intervening, covert psychological magnitude Ψ , unless we have an acceptable theory as to why the output function f_2 in (2) should be precisely the one of simple proportionality, $R = k\Psi$.

For this purpose, moreover, such a theory should presumably be more fully articulated and confirmed than either the rather vague evolutionary argument offered above or the remotely connected argument of simplicity seemingly invoked in the discussion of this matter by Stevens (1964). Given that the overall relation $R = f_3(S)$ is empirically found to be a power function, it may in some sense be simpler (as Stevens implied) to assume that the two component functions, f_1 and f_2 , are themselves both power functions than to assume that the input function is logarithmic (as proposed by Fechner) while the output function is the counteracting exponential. Later, however, Stevens adopted the attractive and plausible hypothesis that magnitude estimation is really just a special case of "cross-modality matching;" and, within that more general framework, the argument based on simplicity changes so as to lose much of its force.

SOME APPARENT LIMITATIONS OF CROSS-MODALITY MATCHING

Partly in order to answer some of the persistent objections to his interpretation of the numerical responses obtained under direct

magnitude estimation, Stevens (1959) introduced and subsequently placed increasing stress on results of experiments in what he called "cross-modality matching." In these experiments the subject is presented, one by one, with stimuli that differ along some sensory continuum just as in magnitude estimation; but, instead of giving a number for each stimulus, the subject adjusts another, variable stimulus along a second sensory continuum until it seems to "match" each presented stimulus. Thus he may adjust a tone of variable intensity until its apparent loudness seems psychologically equivalent to each apparent brightness in a series of lights varying only in intensity.

As in the method of direct magnitude estimation, the subject may be explicitly instructed to maintain proportionality of psychological magnitudes. So, if the second light seems just twice as bright as the first, say, the second tone should be adjusted to sound just twice as loud as the first, and so on. Indeed, Stevens came to hold that magnitude estimation can itself best be regarded as a kind of cross-modality matching (Stevens, 1966b, p. 388). The same can also be claimed for the reverse procedure of "magnitude production," in which a number is given to the subject who then tries to match it by producing an appropriate physical intensity (whether indirectly, by adjusting the attenuator on a tone generator, or directly, by simply singing the desired tone).

The only special features of these latter two varieties of cross-modality matching is that, for one of the two "modalities," a physically given continuum of intensity (or extensity) is replaced by the conventionally established continuum of numbers. In all of these cases, according to Stevens, what the subject really does is to search for the psychological magnitude on one continuum that appears to "match" the psychological magnitude on another continuum. The results tend to substantiate his claim that it makes little difference whether both of the continua are physical or whether one is the learned or conventional "continuum" of numbers.

Indeed the overall consistency of the results that have emerged from these various types of experiments is quite impressive. Specifically, for each pair of continua studied, the intensities chosen on the one continuum have been found to be closely fitted by a power function of the corresponding given intensities on the other con-

tinuum. Moreover, the exponent of this fitted power function is uniformly in close agreement with the value predicted simply by taking the appropriate ratio of the exponents obtained by magnitude estimation for each of the two continua separately (e.g., Stevens, 1959, 1966a). The exponent estimated for each of the sensory continua apparently does represent a real, psychophysical property of that continuum.

However, the basic objection evidently remains that we are still only relating physical magnitudes. Although it is now clearly legitimate to speak of ratios or differences between these physical magnitudes, it is not clear that we have moved any closer to the measurement of the intervening, purely psychological magnitude Ψ .

The model for magnitude estimation, which was formally stated in the earlier Equations (1) and (2), is now replaced by the model for cross-modality matching, which I propose to express in the form

$$\begin{aligned}\Psi &= f(S), \\ \Psi' &= f'(S'), \\ \text{"Match" when } \Psi &= \Psi'.\end{aligned}\tag{4}$$

Here the prime indicates a second sensory continuum (or, in application to magnitude estimation, the "continuum" of conventional number words). The derived relation between empirical variables, which earlier took the stimulus-response form of Equation (3), now takes the stimulus-stimulus form

$$\begin{aligned}S &= f^{-1}\{f'(S')\}, \\ S' &= f'^{-1}\{f(S)\}.\end{aligned}\tag{5}$$

Although this revised model seems more general and is more consonant with Stevens's later view that all psychophysical judgments are based upon matching, a comparison of Equations (3) and (5) reveals that, as far as their empirical implications are concerned, the two models are formally equivalent. In either case we simply end up with a functional relation of the general $X = g(Y)$; and, regardless of whether X is conceived as a response or as a stimulus, X and Y are both observables, so the functional relation g between them is equally susceptible to empirical determination.

More significantly, however, the two models are also alike in that neither provides any empirical basis for deciding among alternative

factorings of the overall function g into its two theoretical components (f_1 and f_2 in the magnitude estimation model, or f and f' in the "matching" model). The basic empirical finding, viz., that the physical intensities that are matched between any two continua are closely describable by a power function, is undisputed. But, contrary to the impression given by Stevens, this finding does not by itself entail that the inner, psychological magnitudes, on the basis of which the subject is assumed to be achieving these matches, are themselves power functions of the external intensities.

As several commentators have noted (e.g., Ekman, 1964; Luce & Galanter, 1963, p. 280; MacKay, 1963; Treisman, 1963), the very same finding could as well be explained by assuming (with Fechner) that the directly-to-be-matched psychological magnitudes all arise, not through a power transformation of the form $\Psi = aS^\beta$, but through a logarithmic transformation of the form $\Psi = a + \beta \log S$. For suppose (with Stevens) that intensities S and S' on two different continua (the unprimed and the primed continuum) result in the desired psychological match, $\Psi = \Psi'$, whenever

$$aS^\beta = a'S'^{\beta'}$$

(where β and β' are fixed parameters, each associated with its own continuum). But this equations is exactly equivalent, mathematically, to the equation

$$\log a + \beta \log S = \log a' + \beta' \log S',$$

which defines a match according to the logarithmic model.³

Stevens himself uniformly applied a logarithmic transformation to the physical intensities, S and S' , before plotting the resulting "matching functions." This is, of course, a perfectly natural and convenient thing to do—particularly since engineers commonly express physical intensities in terms of a logarithmic transformation of energy (i.e., in decibels). However, there are two more reasons for applying such a transformation here. First, the power relation between S and S' is made, in this way, to appear as the graphically simpler linear relation between $\log S$ and $\log S'$. And, second, the variability of the judgments will, in accordance with Weber's law, appear of more uniform size in the resulting log-log plot. (One might even argue that, since the unit of discriminability is approximately constant only on the logarithmic scale, considerations of simplicity

of the overall psychophysical picture—including data on both direct judgment and discriminability—favors the logarithmic over the power model.)

In any case, when magnitude estimation is recast into the general framework of cross-modality matching, it is no longer necessary to invoke an exponential output transformation to counteract a logarithmic input transformation. For now, in place of one input transformation (1) and a separate output transformation (2), we have two symmetrically related input transformations (4). Clearly all such input transformations could be either logarithmic or power functions, and, in view of the above discussion, the previously mentioned argument against the logarithmic alternative on grounds of overall simplicity no longer commands much conviction.

Hypothetical magnitudes Ψ_i that can be reached only through inherently indeterminate functions such as those contained in Equations (3) and (5) have the unsatisfactory status of “nomological danglers” (Feigl, 1958, pp. 382, 428) and ordinarily would suggest a hasty recourse to Ockham’s razor. In the present case, however, to bypass the intervening variable Ψ would be to forfeit the major part of our predictive power. For only by factoring the overall, empirically determined relations (3) or (5) into their two theoretical components does it become possible to predict the overall relation to be found between new combinations of previously studied continua. It is on such grounds that Stevens might argue—as have Hull (1943, pp. 111, 122) and others, including myself (Shepard, 1958b)—for the retention of an intrinsically unobservable variable intervening between the observable stimuli and responses.

Regardless of the ontological status accorded such intervening magnitudes in general, though, they appear to play a curiously indeterminate role in the two models for direct psychophysical judgment considered here. We are left with the following rather awkward state of affairs: In order to provide a simple account for the whole range of cross-modality data, it appears desirable to assume that, corresponding to the external, physical magnitudes, there are internal, psychological magnitudes with definite, quantitative values. But, unless we arbitrarily assume some particular form for the input (or output) functions, nothing can be learned about these values beyond their merely ordinal relations. In such a case it does not

seem wholly justifiable to speak of the "measurement" of a psychological magnitude such as the loudness of a tone or the brightness of light.

FAILURE TO REPRESENT THE RELATIVITY OF PERCEPTION

All of the above objections have taken for granted the traditional psychophysical presupposition that a single stimulus intensity S gives rise, quite by itself, to a quantitatively unique magnitude Ψ — i.e., without reference to any other comparison stimulus. That is, both of the two models considered were models for "absolute" psychophysical judgment. In addition to the objections already raised, I want now to raise the different objection that this traditional presupposition may itself be wrong. For, if an isolated stimulus gives rise to a unique magnitude Ψ , the subject himself seems strangely incapable of making any use of it. The uncertainty and variability of absolute judgments are notorious.

Typically a subject cannot reliably distinguish more than about six or seven *absolute* levels along any one unidimensional continuum (Miller, 1956). This is in striking contrast to a subject's refined sensitivity to *relative* differences between stimuli that differ only slightly, which underlies the hundred or more discriminable steps of Fechner's *jnd* scale. Even in the case of widely separated intensities, a subject may be able to report with considerable reliability and confidence that, with respect to one intensity, the other is greater by half again, by three and one half times, by nearly twenty times, or whatever the case may be.

This superiority of relative judgments is understandable in physiological terms. As we noted at the outset, the absolute level of neural activity depends upon both the intensity of the stimulus and the physiological state of the receptive system. Hence the absolute level of neuronal activity can provide only a very crude and variable indication of the absolute intensity of the external stimulus, and it would not be adaptive for an organism to rely on it for anything more. But, although the absolute rates for two fixed stimuli may thus vary over a wide range, some relationship *between* those rates (such as their difference or ratio) may remain relatively constant (Cornsweet, 1970, pp. 245 ff.). Thus it could well be adaptive for an organism to rely, instead of on the absolute rates of firing, on

something like *relative* rates as valid indicators of what is going on in the external world. In the case of physical measuring instruments, of course, it is elementary that determinations of relative energy are both easier and more accurate than determinations of absolute energy.

It is not surprising, then, that the preferred operations of the psychophysicist typically require the comparison of two or more stimuli. In the method of magnitude estimation, for example, little significance is customarily attached to the number that the subject produces in response to the first stimulus in the series. Although this number may be to some extent influenced by the intensity of the first stimulus, it is typically regarded as a more or less arbitrary "modulus," and the subject will, in fact, accept a quite arbitrary number for this "modulus." It is only with the presentation of the second stimulus that the quantitative value of the subject's responses become of real interest. For at that point he is for the first time constrained by the instruction that, relative to the modulus, the numbers given to succeeding stimuli should be proportional to the magnitudes of the sensations to which they give rise.

A subject can of course tell us something about the intensity of an individual stimulus, if only to assign it to one of a small number of levels (such as "cold," "cool," "warm," and "hot"). However, it is debatable whether a stimulus is ever presented in total isolation, so the judgment may always be at least in part relative. Thus, in judging the brightness of a single spot of light, the subject may compare the spot with the spatially adjacent and temporally coincident surround or with the temporally adjacent and spatially coincident adapting field. We can of course attempt to eliminate all bases for a relative judgment. In the case of brightness, we could reduce the spatial adjacency by substituting a completely uniform field ("*ganzfeld*") for the circumscribed spot, and we could reduce the temporal adjacency by bringing the illumination up to its final physical intensity only gradually. Significantly, however, we then come face to face with the problems of shifting adaptation and, more importantly, with the curious instability, indeterminacy, and even intermittent cessation of visual experience that is characteristic under prolonged *ganzfeld* conditions (Cohen, 1957; Hochberg, Triebel, & Seaman, 1951).

Conditions that would truly force an absolute judgment may be realizable only in *gedanken* experiments. Imagine a line segment that, although it does have a definite physical length (e.g., on the retina), is somehow presented in such a way as to prevent comparison of its visual extent with any reference extent (such as its distance, its width, another visual object, or even the apparent extent of the visual field as a whole). Although the line could presumably still appear to be extended under such hypothetical conditions, it is doubtful whether its extent could be appreciated as a uniquely defined psychological magnitude Ψ . Not only do subjects have difficulty in assessing truly absolute retinal extent within the same eye (Rock, 1975, p. 37), they are often unable even to determine which eye has been stimulated (Pattie, 1935; Pickersgill, 1961). Again, natural selection has favored nervous systems that are primarily tuned to what is "out there" in the external world; and, as in the case of physical measurements, what is out there is most easily and accurately determined by making comparisons.

Those instances in which there does appear to be some basis for genuinely absolute judgment seem to involve the "metathetic" continua such as those of auditory pitch or, possibly, visual hue rather than the intensive or extensive "prothetic" continua considered here. Moreover, even on metathetic continua, relative judgment may typically play a more prominent role than absolute judgment. In the case of pitch, although anyone with a "musical ear" can identify one tone as a third above, or a fifth below another, say, only a very small fraction of the population is able to identify the pitch of a single note absolutely (Siegel, 1972; Ward, 1963). The familiar ability of normal subjects to identify a spectral hue (as red, orange, yellow, green, blue, or violet) might be considered a commonly occurring visual analogue of auditory absolute pitch. But Land's compelling demonstrations show that the experience of a particular color, although it may have a subjectively unique quality, generally depends upon a comparison across a color boundary (Land, 1964, 1966; McCann, 1972). And, again, when the possibility of such comparison is removed under *ganzfeld* conditions, the experienced color tends to desaturate and even to disappear entirely (Hochberg et al., 1951).

Whether or not all perception is relative, it does appear that the perceived magnitude of a stimulus on a prothetic continuum, at least, is generally largely, if not totally, relative to some comparison magnitude on that same continuum. But, if so, any model that deals with but a single stimulus at a time on each continuum may already impose the wrong sort of structure.

Reinterpretation via a "Relation Theory" of Psychophysical Judgment

It was the recognition of the importance of the relativity of perception that led me, over ten years ago, to investigate the consequences of this relativity for the fundamental indeterminacy of "direct" psychophysical scaling demonstrated in the preceding sections of this paper. Toward this end I proposed, in place of either of the two models for absolute psychophysical judgment given in Equations (1-3) or (4-5), a rather different model, which Krantz, in his well-formulated development of it, has named the "relation theory" (Krantz, 1972a).⁴ The following sections are devoted to the examination of this alternative theory, its justification, and its consequences for the problem of "direct" psychophysical measurement.

RECONSIDERATION OF THE FUNDAMENTAL MATCHING OPERATIONS OF PSYCHOPHYSICS

Stevens's suggestion that all psychophysical judgment amounts to the matching or equating of external, physically specifiable things with respect to their inner, psychological effects has considerable appeal. As Stevens pointed out, this suggestion uniformly places the subject in the conceptually simple role of a sort of "null instrument." Or, in the spirit of the cybernetic analysis of control through feedback (cf. MacKay, 1963; Powers, 1973), to produce a response (whether by squeezing a hand dynamometer in "magnitude production" or by finding an appropriate number in "magnitude estimation") is really to instate a certain stimulus; namely, the one that results in a match of the corresponding internal magnitude Ψ' to the appropriate comparison magnitude Ψ . Thus, when the subject in a magnitude estimation experiment gives a particular number, say "45," to a particular intensity of tone, he is merely telling us which number matches the given tone with respect to the two cor-

responding internal magnitudes, Ψ (for the tone) and Ψ' (for the number).

However, according to the preceding argument for the relativity of perception, even if every psychophysical judgment amounts to the establishment of a match between two things, the two things that are thus said to be psychologically equivalent may not be two individual stimuli but, at least, two *pairs* of stimuli.

In some situations this structure may be imposed quite explicitly, as when the subject is given a pair (S_i, S_j) and is asked to adjust a variable stimulus S_x in a second pair (S_k, S_x) until the two pairs become in some way psychologically equivalent. If the physical intensity of S_j is just twice the physical intensity of S_i , this might happen when, for example, the physical intensity of S_x is adjusted to just twice the physical intensity of S_k . The same sort of operation can of course be performed, too, when there are only three stimuli—as when the pairs to be compared are, in effect, (S_i, S_j) and (S_j, S_x) .

In other situations, such as cross-modality matching, the composition of the pairs may not be explicitly specified. Suppose a subject is required to adjust the loudness of a variable tone S'_x until it matches the brightness of a fixed spot of light S_i (where the prime indicates that the variable stimulus is on a different continuum). The subject may not be able to do this directly, but only derivatively—by comparison of the pairs (S_0, S_i) and (S'_0, S'_x) , where S_0 and S'_0 are corresponding implicit reference stimuli for the two continua (perhaps the initial or background levels of visual and auditory stimulation respectively).

The response to the first stimulus in a magnitude estimation task could be generated in a similar manner, provided that numbers are treated like stimuli on any purely sensory continuum. The variable and often seemingly arbitrary character of this first response could still be explained in terms of the subject's uncertainty about the unspecified reference levels S_0 and (for the number continuum) S'_0 . (Then too, this uncertainty might be subject to amplification owing to the generally great distance of the presented stimulus S_i from the chosen reference stimulus S_0 .) So, again, it is only with the presentation of the second, explicitly constraining stimulus that we can expect an orderly pattern to emerge in the subject's responses.

Now it might seem that, with the presentation of the third and subsequent stimuli, our analysis must rapidly increase in complexity. It might even appear that we should have to specify whether each successive stimulus is related just to the immediately preceding stimulus or to some weighted combination of those presented most recently—a specification that could easily involve us in unresolved issues concerning temporal information processing, mechanisms of short- and long-term memory, attention, and the like. Fortunately, according to the relation theory, an understanding of the logical basis of psychophysical scale construction can be gained without making such further specifications. For it follows from that theory that all of the information is internally consistent (“transitive”) in such a way that, except for distortions introduced by memory itself, a subject’s judgment should be the same regardless of which preceding stimuli are taken as a basis for the judgment.⁵

EQUIVALENCE CLASSES OF PAIRS OF STIMULI

Instead of starting, as in classical psychophysics, with a mapping of S into Ψ , we begin, in the relation theory, with a mapping of the Cartesian product set $S \times S$ into Ψ . That is, the results of the fundamental matching operations define equivalence classes of all pairs (S_i, S_j) that are judged to be psychologically equivalent and that, for this reason, are assigned to the same subjective magnitude Ψ . The psychophysical data also enable us to establish an ordering of these equivalence classes. For, if two pairs (S_i, S_j) and (S_b, S_k) are clearly *not* equivalent, the subject can tell us, not only that they are not equivalent, but also in what direction they depart from equivalence. He can say whether the contrast presented by the pair (S_i, S_j) is greater or less than the contrast presented by the pair (S_b, S_k) . Since such judgments are generally consistent (i.e., transitive) for all except marginally equivalent pairs, they define an ordering on the set of psychological magnitudes, Ψ .

In order to go further, we need to specify something about the structure of these equivalence classes or, in other words, about the nature of the function $\Psi(S_i, S_j)$, mapping each pair, S_i and S_j , into an element of the ordered set Ψ . For this purpose we must, for any given continuum, adopt a particular measure of the physical magnitude of any stimulus on that continuum. What we want is some

measure of the *amount* of the stimulus, so that the physical measure is properly additive if we combine stimuli. For intensive continua, the most reasonable measure seems to be the traditional one of physical energy.

The psychophysical data of Stevens and his students then show that, to a first approximation, *it is equal ratios of physical magnitude that are psychologically equivalent*. This amounts to saying that, if we increase the physical magnitudes of any two stimuli by the same constant factor k , the psychological relation between them remains invariant; in functional notation,

$$\Psi(S_i, S_j) = \Psi(kS_i, kS_j), \tag{6}$$

provided, of course, that S_i, S_j , and k are all greater than zero.

This is, moreover, a rule with considerable adaptive utility. As Plateau noted in 1872, an object with surfaces of characteristic reflectances could then be recognized as the same object under widely different levels of illumination (Herrnstein & Boring, 1966, pp. 75-79). Likewise, figures composed of lines of different lengths could be recognized as the same figure when viewed from different distances (cf. Rock & Ebenholtz, 1959); speech sounds could be recognized whether loud or soft; music would sound much the same whether a record is played at 33 or 45 rpm; and, perhaps similarly, speech could be understood whether uttered by a young child or an adult male.

The implications of Equation (6) are perhaps easier to see if we display the iso- Ψ contours, graphically, in the $S_i \times S_j$ space, as shown in Figure 1. What (6) requires is that these contours all be straight lines radiating from the origin. (The origin itself as well as narrow regions all along both zero-intensity axes should, as indicated by the broken lines, be excluded by proper ancillary conditions on this equation.)

Each of these linear contours is defined by the similarity relation $S_j = c \cdot S_i$ or, in other words, by the requirement that the ratio S_j/S_i is a constant, c . Now (6) also requires that $\Psi(S_i, S_i) = \Psi(S_j, S_j)$ for all i and j . Clearly then, the function Ψ must have the form

$$\Psi(S_i, S_j) = f\left(\frac{S_j}{S_i}\right), \tag{7}$$

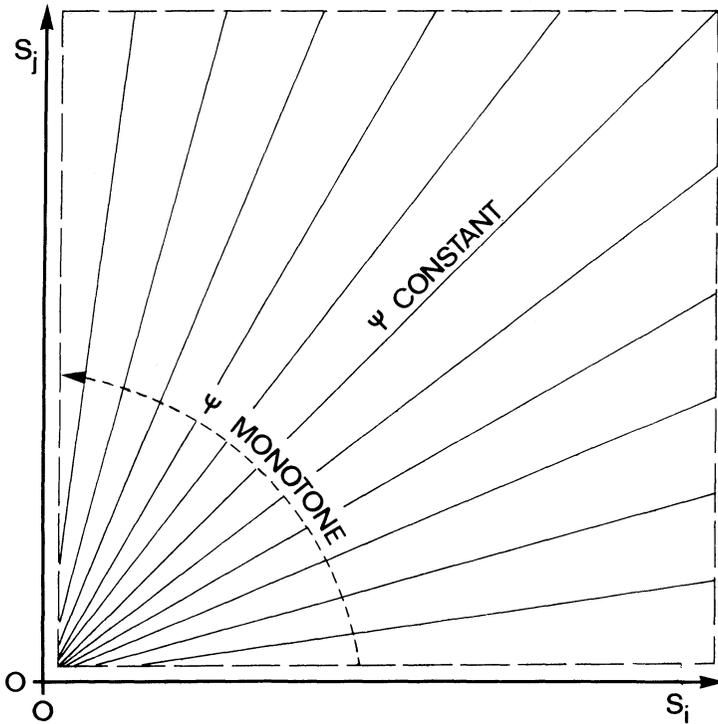


Figure 1. Loci of points corresponding to pairs of physical magnitudes, S_i and S_j , on the same sensory continuum that produce psychological contrasts of the same magnitude Ψ .

where f is any function. Whereas in the classical formulation Ψ is the subjective magnitude corresponding to the intensity of a single stimulus, in the present formulation Ψ is the subjective magnitude corresponding to the ratio of two stimuli.

In addition, we noted that we can establish an ordering on these subjective magnitudes. Within any one continuum, this ordering is determined by the requirement that

$$\Psi(S_b, S_k) > \Psi(S_i, S_j) \tag{8}$$

whenever $S_k/S_b > S_j/S_i$.

From this further condition it then immediately follows that the function f in (7) is monotone increasing, as indicated in Figure 1.

Now, if we are given any two intensities S_1 and S_2 on the same continuum, we can construct a sequence, $S_1, S_2, S_3, S_4, \dots$, such that the psychological magnitude Ψ is the same for all adjacent pairs; i.e., so that for any j

$$\Psi(S_j, S_{j+1}) = \Psi(S_{j+1}, S_{j+2}) = \text{const.} \tag{9}$$

But, according to (8), this psychological equivalence entails a physical relation in which the ratios between successive intensities are all the same; i.e., in which

$$\frac{S_2}{S_1} = \frac{S_3}{S_2} = \frac{S_4}{S_3} = \dots = \frac{S_i}{S_{i-1}} = \text{const.}$$

And from this it follows that the intensity S must be expressible in terms of the two originally given intensities, S_1 and S_2 , as follows:

$$S_i = \left(\frac{S_2}{S_1}\right)^{i-1} \cdot S_1 \tag{10}$$

In short, the assumption (6) leads directly to the conclusion that, if the psychological relations between successive stimuli are to be constant, the stimuli must form a geometric series: viz.,

$$S_1, cS_1, c^2S_1, c^3S_1, \dots,$$

where c is the ratio between the two initially given intensities.⁶

RELATIONS AMONG DIFFERENT CONTINUA

The results obtained so far have shown how we might construct a sequence of stimuli such that the psychological effect is the same for any pair of stimuli separated by the same number of steps along the sequence. In this sense, of course, we might be said to have provided a basis for a scaling of the physical stimuli with respect to their psychological effects. We have not, however, provided any basis for a scaling of those psychological effects themselves; and surely, it would not be correct to say that we are now in a position to measure any internal, psychological magnitudes. Stevens’s work suggests that if a basis for such a scaling or such measurements is to be found, it is probably to be found in the psychological relations among different sensory continua. Let us turn, therefore, to a consideration of “cross-modality” matching which, as we shall see, does place one further constraint on our model for relative psychophysics.

Suppose, then, that we have two sensory continua, the primed and the unprimed (corresponding, for example, to continua of auditory and visual intensity). If our psychological matching operations can be extended from one continuum to the other, they will define a mapping of the Cartesian product sets $S \times S$ and $S' \times S'$ into either the same ordered set, Ψ , or equivalently, into two ordered sets, Ψ and Ψ' , with a uniquely defined one-to-one correspondence between them. In order to be consistent with the formulation of the matching model, we shall adopt the latter interpretation.

It then follows that, if we are given any pair (S_1, S_2) on the unprimed continuum, we can not only construct a sequence S_1, S_2, S_3, \dots on that same continuum satisfying (9), we can also construct a sequence S'_1, S'_2, S'_3, \dots on the primed continuum such that

$$\Psi'(S'_i, S'_{i+1}) = \Psi(S_i, S_{i+1}) = \text{const.} \quad (11)$$

The first stimulus, S'_1 , on the primed continuum may be arbitrarily supplied by the experimenter or it may be left to the subject to choose either arbitrarily or, possibly, by resorting to comparisons with some implicit reference stimuli, S_0 and S'_0 , to achieve at least a rough match between the pairs (S_0, S_1) and (S'_0, S'_1) . (We do not suppose that a pair such as (S_1, S'_1) produces a unique psychological magnitude directly when the two intensities are on different continua.) In any case, once S'_1 has been fixed, the remainder of the sequence of stimuli needed to satisfy the matching condition (11) on the primed continuum will be rigidly determined. Indeed, just as in the case of the unprimed continuum, this sequence will form a geometric series:

$$S'_i = \left(\frac{S'_2}{S'_1} \right)^{i-1} \cdot S'_1. \quad (12)$$

The two geometric sequences and the correspondence between them are illustrated graphically in Figure 2. If the first two stimuli on the unprimed continuum had been at the intensities of S_2 and S_4 (instead of at the intensities of S_1 and S_2), while S'_1 was still retained as the first stimulus on the primed continuum, then the second corresponding stimulus on the primed continuum would have been S'_3 and the new correspondence would be as shown by the dashed lines.

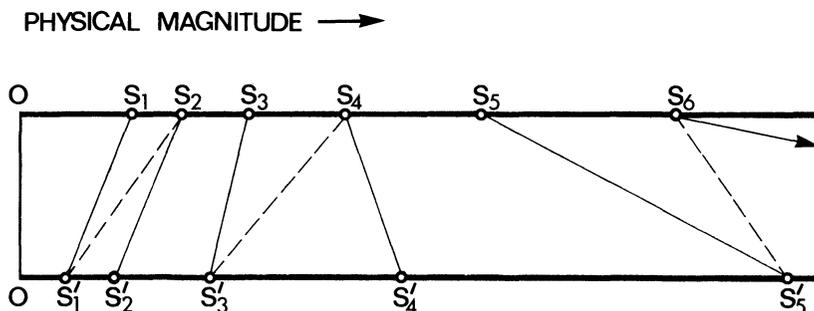


Figure 2. Psychological correspondences between physical magnitudes on two different continua, such as those for brightness and loudness.

So far, we have two geometric series of stimuli: one (12) on the primed continuum and one (10) on the unprimed continuum. From the assumptions, two things follow. First, within either series, stimuli separated by the same number of steps constitute psychologically equivalent pairs. And, second, between the two series, pairs separated by one step in one series are psychologically equivalent to pairs separated by one step in the other series, as required by (11). However, for $n > 1$, it does not yet follow that pairs separated by n steps in one series (although equivalent with each other) are also equivalent with pairs separated by n steps in the other series. This further correspondence between the two series apparently requires the explicit introduction of an additional assumption.

My former colleague David Cross, who first made me aware of the importance of this additional assumption, proposed that it be called the assumption of "transitivity" of equivalence relations. It takes the form

$$\begin{aligned}
 &\text{if } \Psi'(S'_i, S'_j) = \Psi(S_i, S_j) \\
 &\text{and } \Psi'(S'_j, S'_k) = \Psi(S_j, S_k), \\
 &\text{then } \Psi'(S'_i, S'_k) = \Psi(S_i, S_k).
 \end{aligned}
 \tag{13}$$

This assumption seems to be well supported by the empirical results obtained by Stevens and others from experiments on cross-modality matching. As we shall see, it implies that the function f in (7) cannot have an arbitrarily different form for each continuum.

What, then, must the relation be between the functions f and f' for two continua? Notice that, in the two sequences that have been matched on the primed and unprimed continua, regardless of the values of the first four physical magnitudes S_1, S_2, S'_1 , and S'_2 , (always finite, positive numbers), there will exist a unique number p given by

$$p = \log \left(\frac{S'_2}{S'_1} \right) / \log \left(\frac{S_2}{S_1} \right).$$

Hence, the relation between the ratios of the first two intensities on the two continua can always be expressed in the form

$$\frac{S'_2}{S'_1} = \left(\frac{S_2}{S_1} \right)^p \quad (14)$$

If, now, we substitute (14) into (12), we obtain

$$S'_i = \left(\frac{S_2}{S_1} \right)^{p(i-1)} \cdot S'_1.$$

But according to (10),

$$\left(\frac{S_2}{S_1} \right)^{(i-1)} = \frac{S_i}{S_1},$$

which can be substituted into the preceding equation to obtain

$$S'_i = \left(\frac{S_i}{S_1} \right)^p \cdot S'_1$$

or, after rearranging terms,

$$\frac{S_i^p}{S'_i} = \frac{S_1^p}{S'_1}.$$

This, of course, must hold for any j as well as for any i , so it follows more generally that

$$\frac{S_i^p}{S'_i} = \frac{S_j^p}{S'_j} = \frac{S_1^p}{S'_1}.$$

A final rearrangement of terms in the left and middle ratios then yields

$$\frac{S'_i}{S_i} = \left(\frac{S_j}{S_i} \right)^p \quad (15)$$

for any i and j .

Moreover, the exponent p depends only upon the two continua themselves, and not at all upon the particular stimuli S_1 , S_2 , and S'_1 arbitrarily chosen to construct the two matching sequences on these continua. Thus, in Figure 2, the alternative correspondence between stimuli indicated by the broken lines still leads to the very same exponent p . Apparently, then, the inherent differences in the way two prothetic continua operate psychologically can be fully accommodated simply by expressing the relation between the two continua in terms of a power transformation with a uniquely defined exponent or power p as indicated in (15).

Indeed, the relations among all such continua can be most parsimoniously explained simply by associating with each of the individual sensory continua, say, S , S' , S'' , etc., a characteristic power, p , p' , p'' , etc., such that the exponent characterizing the relation between any two of these continua is given simply by the appropriate ratio of the two powers associated with the continua compared. A consequence is that, while the exponent relating any two continua is determined completely (i.e., up to the identity transformation), the exponents associated with the individual continua themselves are determined only up to multiplication by an arbitrary constant. Hence in (15) the exponents associated with the primed and unprimed ratios could as well have been q and qp (with any arbitrary q), instead of simply 1 and p .

In practice, of course, this arbitrary factor q can be fixed for all continua by adopting the convention that the exponent characterizing some particular or preferred continuum (such as the continuum of visual length or that of the number responses in direct magnitude estimation) shall be unity. The empirical results of the cross-continuum matching experiments, together with this convention, then suffice to determine the exponents for all other continua uniquely (to within experimental error).

Apparently, then, the fundamental relation of relative psychophysics can be written in the final, more explicit form

$$\Psi(S_i, S_j) = g \left\{ \left(\frac{S_j}{S_i} \right)^p \right\}, \quad (16)$$

where, again, g is any monotone increasing function. Now, however, in order to ensure satisfaction of the "transitivity" condition (13), the function g , whatever its form, must be regarded as having the *same* form for all continua. The difference between the psychophysical functions, f and f' , for any two different continua is thus entirely absorbed into the power function indicated by the exponent p .

THE IRREDUCIBLE

INDETERMINACY OF THE PSYCHOPHYSICAL FUNCTION

It is clear that, if we are given any sequence of physical magnitudes S_a, S_b, S_c, \dots on one continuum, and any one intensity S'_a on a second continuum, we can construct a sequence S'_a, S'_b, S'_c, \dots on the second continuum such that, for any i and j

$$g \left\{ \left(\frac{S_j}{S_i} \right)^p \right\} = g \left\{ \left(\frac{S'_j}{S'_i} \right)^{p'} \right\}, \quad (17)$$

where p and p' are the powers associated with the two continua in question. Clearly, too, the form of the monotone increasing function g is totally irrelevant to the outcome. (Application of g^{-1} to both sides has no empirical consequences.)

In order to account for the principal results that Stevens and his followers have found for cross-modality matching (including magnitude estimation and magnitude production), then, we need not assume that there is, corresponding to a given intensity S_i , any uniquely defined subject magnitude $\Psi(S_i)$. Instead we need merely assume that for any pair of intensities S_i and S_j there is some subjective magnitude

$$g \left\{ \left(\frac{S_j}{S_i} \right)^p \right\},$$

where g is some fixed, monotone increasing function whose form is not otherwise constrained by the psychophysical data.

We could, of course, assume that

$$g(x) = ax^\beta \tag{18}$$

(where, as an even more special case, we could take $a = \beta = 1$). Then we would have

$$\Psi(S_i, S_j) = a \left(\frac{S_j}{S_i} \right)^{\beta p}, \tag{19}$$

and none of the previous conclusions would be altered by this specialization. If, then, we take $S_i = S_0$, this amounts to the power law adopted by Stevens.

On the other hand, we could as well assume that

$$g(x) = a + \beta \log x, \tag{20}$$

in which case we would find that

$$\Psi(S_i, S_j) = a + \beta p (\log S_j - \log S_i). \tag{21}$$

Again nothing would be changed, although in this case we see that, with respect to the new physical variable $\log S$, Ψ can be regarded as a difference rather than as a ratio of physical magnitudes.

ALTERNATIVE DERIVATION USING FUNCTIONAL EQUATIONS

By taking advantage of known solutions to certain functional equations, the above-demonstrated irreducible indeterminacy of the psychophysical function can be established by an alternative, more elegant derivation, subsequently proposed to me by David Cross. Strictly, in addition to the Invariance Assumption (6), stated separately for the primed as well as the unprimed continuum, and the Transitivity Assumption (13), we need to make an implicit Comparability Assumption explicit, as follows: For any S_i and S_j in the unprimed continuum and any S'_i in the primed, there exists a S'_j in the primed continuum such that

$$\Psi(S_i, S_j) = \Psi'(S'_i, S'_j). \tag{22}$$

As we already noted in deriving Equation (7) from (6), the most

general solutions to the functional equations in the Invariance Assumption are

$$\Psi(S_i, S_j) = f\left(\frac{S_j}{S_i}\right) \text{ and } \Psi'(S'_i, S'_j) = g\left(\frac{S'_j}{S'_i}\right),$$

where f and g are arbitrary functions.

However, by the Comparability Assumption (22),

$$f\left(\frac{S_j}{S_i}\right) = g\left(\frac{S'_j}{S'_i}\right).$$

By applying g^{-1} to both sides, we obtain

$$\frac{S'_j}{S'_i} = b\left(\frac{S_j}{S_i}\right), \text{ where } b = g^{-1}f.$$

Then, by the Transitivity Assumption (13),

$$\frac{S'_k}{S'_j} = b\left(\frac{S_k}{S_j}\right) \text{ and } \frac{S'_k}{S'_i} = b\left(\frac{S_k}{S_i}\right).$$

But

$$\frac{S'_k}{S'_i} = \frac{S'_k}{S'_j} \cdot \frac{S'_j}{S'_i} = b\left(\frac{S_k}{S_j}\right) \cdot b\left(\frac{S_j}{S_i}\right).$$

Thus

$$b\left(\frac{S_k}{S_i}\right) = b\left(\frac{S_k}{S_j}\right) \cdot b\left(\frac{S_j}{S_i}\right). \tag{23}$$

For some $a > 0$ and $b > 0$,

$$S_j = aS_i, S_k = bS_i, \text{ and hence } S_k = abS_i.$$

Substituting into (23), we have

$$b(a \cdot b) = b(a) \cdot b(b).$$

This classical functional equation of Cauchy has, as its most general continuous solution, the power form

$$b(t) = t^p$$

Hence, for any i and j , if

$$f\left(\frac{S_j}{S_i}\right) = g\left(\frac{S'_j}{S'_i}\right),$$

the forms of f and g , instead of being arbitrary, are constrained by the composition

$$g^{-1}f\left(\frac{S_j}{S_i}\right) = \left(\frac{S_j}{S_i}\right)^p. \tag{24}$$

Cross noted that two classes of functions that satisfy the functional composition (24) are the class of logarithmic functions

$$f(x) = k \ln x, \quad g(y) = c \ln y,$$

and the class of power functions

$$f(x) = x^a, \quad g(y) = y^b.$$

On further consideration, however, it appeared to me that there is a more general class of functions (which subsumes the above two classes) that also satisfies the same functional composition (24); viz.,

$$f(x) = h(x^a) \quad \text{and} \quad g(y) = h(y^{a/p}),$$

where h is any monotone increasing function. For if

$$g(x) = f_1[f_2(x)],$$

then

$$g^{-1}(x) = f_2^{-1}[f_1^{-1}(x)],$$

so

$$g^{-1}\left[f\left(\frac{x'}{x}\right)\right] = h^{-1}\left[h\left\{\left(\frac{x'}{x}\right)^a\right\}\right]^{p/a} = \left[\left(\frac{x'}{x}\right)^a\right]^{p/a} = \left(\frac{x'}{x}\right)^p.$$

Again, therefore, I conclude that the functions f and g can be of any monotone increasing form, although (owing to transitivity) they must always be of the *same* form for all continua. As before, whether these functions are taken, for all continua, to be power functions, log functions, or functions of some other form must be decided on the basis of considerations other than the results of magnitude estimation or cross-modality matching.

Discussion and Conclusions

The relation theory has led to the conclusion that the operations of magnitude estimation and cross-modality matching are not sufficient to determine anything beyond the ordinal relations among the psychological magnitudes, Ψ_i . I want to conclude by explicitly setting down what I take to be the implications of this result for perennial questions concerning (a) what it is that the psychophysicist measures, (b) what the status is of the so-called psychophysical law, and (c) what type of scale it is that is constructed in psychophysical scaling. But first some clarification is in order concerning what appears to be a possible dependence of the relation theory upon implicit absolute psychophysical judgments, and concerning the extent to which the present conclusions are dependent upon the relation theory.

RELATIVE VERSUS ABSOLUTE
PERCEPTION, AND THE ROLE OF MEMORY

In the preceding analysis, relative psychophysical judgment has been described as if the resulting psychological magnitude, say $\Psi(S_i, S_j)$, is something that arises from the comparison of two stimuli, S_i and S_j , that are both available simultaneously. Actually the stimuli within any one continuum are usually presented successively. In most cases, then, it is not entirely correct to say that the two stimuli S_i and S_j are compared directly. At best, each presented stimulus in these cases can be compared only with some sort of memory trace of any preceding stimulus.

This, however, raises a perplexing conceptual problem for the notion that a well-defined psychological magnitude, such as Ψ , arises only upon the presentation of at least two stimuli from any one continuum. For surely, if $\Psi(S_i, S_j)$ assumes a well-defined value even though S_i was removed before the presentation of S_j , then the trace of S_i alone must have had a well-defined value. In other words, how could the ratio S_j/S_i , upon which $\Psi(S_i, S_j)$ depends, be available to the subject unless definite, quantitative information about S_i was somehow preserved within the subject? But, if such quantitative information about individual stimuli is thus represented within the subject, in what sense can it properly be claimed that a definite psychological magnitude Ψ comes into being only upon comparison with a second stimulus?

Indeed the same problem remains even when the two stimuli are presented simultaneously. Even if they are not separated in time, they are nevertheless separated in some other way (e.g., in space). But the two external stimuli are not themselves brought together within the subject's nervous system; what is brought together can only be some internal representation of these two stimuli. Again, if the result of this comparison depends upon the ratio of the two physical magnitudes, then the quantitative information necessary to determine this ratio must be contained in the two internal representations—even before they are brought together.

What, then, is the explanation for the fact that relative judgments are more stable and precise than absolute judgments? Even though there must indeed be some quantitative representation of each stimulus separately, I have already argued that these separate representations will necessarily be highly dependent upon such external variables as distance and condition of illumination that are not inherent in the perceived object itself. Moreover, these separate representations are likely to vary widely with the internal state of the organism. It seems reasonable to suppose that changes in internal state (like variations in external conditions) tend to affect the neurophysiological encoding of magnitudes in a similar way. Rate of neural firing, for example, might be affected additively or multiplicatively. If so, it would be only when two such separate representations are brought together and a difference or ratio formed that a quantitative representation would be obtained that has a suitably invariant relation to the external stimulus.

Such a supposition would explain the well-known fact that even relative comparisons become less stable and precise as the two stimuli are separated in time, space, or along some other dimensions (e.g., wavelength for colors to be matched in brightness, or frequency for tones to be matched in loudness). For, as two stimuli are separated in any way, it becomes more probable that the local internal states prevailing in the relevant nervous centers will differ significantly and thus fail to be canceled out in the subsequent computation of the ratio or difference. This is particularly clear in the case of a separation in time and seems, therefore, to provide a plausible account of the well-established decay in recognition memory with time (cf. Shepard, 1958a; Wickelgren, 1969).

Under favorable conditions, in which the separation is not too great, on the other hand, the computed ratio or difference will be relatively independent of the perturbing physiological parameters since these will affect both components entering into the computation alike. It would not be surprising, then, if natural selection had favored an organization of the brain such that these purified ratios or differences are readily connectable to voluntary responses (including verbal reports), whereas the contaminated raw components of these ratios or differences are kept relatively inaccessible in order to decrease the probability of an inappropriate response to the external world.

Nevertheless, since there must be some sort of quantitative, internal representation or magnitude for each individual stimulus even before any comparison takes place, we cannot dismiss the possibility that subjects may sometimes be able to report truly absolute psychophysical judgments. The only points that I wish to emphasize about such absolute judgments are the following two: First, according to the arguments made just above (and earlier in this paper), such judgments are likely to be less reliable than relative judgments. And second, even under conditions in which such judgments are sufficiently reliable, the analyses presented here—particularly following Equations (3) and (5)—and elsewhere by other commentators indicate that the constraints provided by absolute judgments are no greater than those provided by relative judgments. Our basic conclusion as to the quantitative indeterminacy of the psychological magnitudes Ψ_i does not depend, then, upon the adoption of the relation theory. What, then, must our answers be to the long-debated questions concerning psychophysical “measurements,” “laws,” and “scales”?

WHAT DOES THE PSYCHOPHYSICIST MEASURE?

Usually when one speaks of measuring something, one has in mind the assignment of a number on a numerical (i.e., interval or ratio) scale. Having measured some objects in this sense, one can legitimately report such facts as that one object is twice as long as another, or is equal in weight to the sum of two others, and so on. Naturally, then, when someone speaks of measuring a sensation, it suggests that the sensation has likewise been fixed on a numerical

scale, and that one can thereby determine whether one sensation is twice another, is equal to the sum of two others, and so on.

According to the analysis presented above, however, the operations of magnitude estimation and cross-modality matching upon which Stevens proposed to base psychophysical measurement do not determine any more than an ordinal structure on the psychological magnitudes Ψ_i . So, although the subject himself can tell us that one such inner magnitude is greater than another, the psychophysical operations that we have considered are powerless to tell us anything further about *how much* greater the one is than the other. My conclusion—like those of Krantz (1972a, b), Luce (1972), and Savage (1970)—is that these operations do not in themselves permit us to measure inner sensations in any quantitative sense.

What the psychophysical operations of magnitude estimation and cross-modality matching *do* enable one to measure on a ratio scale is the exponent p that characterizes a given (prothetic) sensory continuum for a given subject or population of subjects. In short, what the psychophysicist measures is an important constant governing how a subject transduces any stimulus from a particular sensory continuum, such as the continuum of lights varying in intensity, tones varying in amplitude, or lines varying in length. He does not measure the magnitude of any one inner subjective sensation produced by one stimulus or pair of stimuli as opposed to another along any such continuum.

WHAT IS THE STATUS OF THE PSYCHOPHYSICAL LAW?

Although it is possible to determine, for each prothetic continuum, the exponent p that characterizes that particular continuum uniquely (up to multiplication of all such exponents by an arbitrary constant), the psychophysical *relations* between physical magnitudes and the postulated inner psychological magnitudes Ψ_i for all such continua are seen to contain the same indeterminate function g . Stevens's claims that the psychophysical law is a power law appear to be implicitly based on the argument that g should be taken, on grounds of simplicity, to be a power function—indeed the identity function. But such an argument seems to me to lack the richness of supportive interconnections with the rest of the sciences of sensation, perception, cognition, and neurophysiology to inspire conviction.

Indeed, a consideration of the neurophysiology of sensory perception suggests that the notion that a physical stimulus gives rise to some *one* unique inner magnitude Ψ_i is rather simplistic. Such a notion seems to be based on the tacit assumption that there is some *one* stage, during the propagation of the sensory signal through the nervous system, at which a coded representation is displayed before some "homunculus" or "ghost in the machine." Instead I find it more satisfactory to suppose that this propagation proceeds through a whole series of transformations. The results of different ones of these transformations may well correspond to different forms of the function g .

Such a view is consonant, also, with evolutionary considerations. For, if it is adaptive for an organism to be able to learn new responses without also having to master, each time, some new and nonlinear function *de novo*, then it will also be useful for the organism to have, for each important physical variable, a choice of representations corresponding to different, generally useful transformations (linear, logarithmic, etc.). Much as in the "pandemonium" model of Selfridge (1958), then, the process of learning would be one in which the response becomes attached to the most appropriate (i.e., simply related) of these available alternative representations. Even if an elucidation of the psychophysical problem along these lines has no other virtue, it at least accounts for the troublesome fact that diverse psychophysical procedures often lead to equally diverse results!

WHAT TYPE OF SCALE IS CONSTRUCTED IN PSYCHOPHYSICAL SCALING?

In Stevens's celebrated classification scheme, scales are assigned to a type on the basis of the group of transformations under which the empirically significant properties of the scale remain invariant (Stevens, 1946, 1951). The most commonly considered types are the *ordinal scale*, in which only order is significant, so any monotone transformation is permissible; the *interval scale*, in which the equivalence or nonequivalence of differences (i.e., intervals) is significant, so the monotone transformation must also be linear; and the *ratio scale*, in which the equivalence or nonequivalence of ratios (as well as differences) is significant, so the linear transformation must be restricted to a similarity (i.e., to multiplication by a constant). When

they can be constructed, ratio scales are preferable to interval scales, which in turn are preferable to merely ordinal scales. The reason is that the former types provide greater quantitative leverage. Indeed, since determination on a merely ordinal scale really provides only qualitative rather than quantitative information, it seems to be stretching the usual meaning of the word "measurement" to apply it to the ordinal case at all.

Whereas Stevens claimed that his psychophysical operations permitted measurement on a ratio scale, this appears to be the case only for the measurement of the parameter p governing each prothetic continuum—not for the measurement of the magnitude of a sensation within any continuum. There is, however, another possible way of construing Stevens's contention concerning scale type. Although we cannot obtain a quantitative determination of the psychological magnitudes themselves, we nevertheless can use the empirically determined equivalences among these magnitudes to construct a spacing or "scale" that does have some quantitatively unique properties *with respect to a particular, designated physical variable*. Thus, if we take energy as our fundamental measure of physical magnitude on a certain prothetic continuum, the psychological equivalences that are determined by cross-modality matching lead to a geometric spacing on that physical variable. Now this spacing is unique (on that variable) up to a power transformation; a geometric series remains geometric after, and only after, all terms are multiplied by a constant and/or raised to a power. Since more general, monotonic transformations destroy this property, the result is stronger than a merely ordinal scale.

It does not, however, correspond to either an interval or a ratio scale in Stevens's sense. Rather, it appears to amount to a different type of scale based upon the equivalence of ratios *without* the prior equivalence of differences. It is somewhat curious, then, that with regard to this new type of scale, later recognized by Stevens and dubbed the "logarithmic interval type,"⁷ Stevens stated that "it has thus far proved empirically useless" (1957, p. 176). On the one hand, he was clearly aware that, in order to convert such a logarithmic interval scale into a full ratio scale, we must be able either (a) to equate differences or else (b) to determine the numerical values—rather than just equivalences—of ratios. But, on the other hand,

he seems not to have made clear how his psychophysical operations provide a satisfactory basis for doing either of these things.

Whatever the type of such a scale, it is not strictly a scale that permits the quantitative measurement of psychological magnitudes. If there are unique psychological magnitudes, these could, as we saw, be related to this scale by any monotonic function. Furthermore, any measurement performed on such a scale would be "derived" rather than "fundamental" measurement (Suppes & Zinnes, 1963), since it depends upon the existence of a previously established physical scale; viz., the ratio scale of energy. Such a measurement is arbitrary in the sense that a change to a different physical variable (say number of decibels, which is logarithmically related to energy) can induce a nonadmissible transformation in the derived scale (the geometric series can become arithmetic).

CONCLUDING REMARKS

The reinterpretation of the results of direct psychophysical judgment in terms of the relation theory seems to offer several advantages. It takes account of the well-established superiority of relative over absolute judgments. It is consonant with evolutionary arguments for the selective advantage of a perceptual system that responds principally to ratios of intensities and extensities. It suggests a general neurophysiological basis for a variety of phenomena of comparative judgment and memory. And, as Krantz (1972) has noted, it leads to empirically testable consequences beyond those directly suggested by the traditional theories of absolute psychophysical judgment. The reinterpretation made possible by the relation theory does not, however, provide a way of circumventing the fundamental indeterminacy of the implications of magnitude estimation and cross-modality matching concerning the subjective magnitudes of sensations.

I conclude that, if we are to pin down any of what may well be several monotonically related types of internal representations of the magnitude of a stimulus quantitatively, we are going to have to move outside the circumscribed system of relationships provided by these "direct" psychophysical operations. Possibly depending on the type of internal representation, we may find the necessary additional relationships in the neurophysiological results of single-

cell recording (Kiang, 1965; Luce & Green, 1972; Mountcastle, Poggio, & Werner, 1963; Perkel & Bullock, 1968; Rushton, 1961); in the behavioral results of experiments on stimulus generalization (Shepard, 1965), discrimination, and disjunctive reaction time (Curtis, et al., 1973; Falmagne, 1971; Shepard, et al., 1975); or in the cognitive results of experiments on the mental combining of perceived magnitudes (Anderson, 1970; Birnbaum & Veit, 1974; Falmagne, 1976; Levelt, Riemersma, & Bunt, 1972; Luce & Tukey, 1964; Sternberg, 1966). According to the theoretical analysis I have presented, the empirical findings of Stevens imply that the internal representations of sensory magnitude that are compared with each other in any given task are all related to physical magnitude by the same function, g . This result must, I believe, have some empirical significance. It encourages me, in any case, to believe that the search for further relationships at the neurophysiological, behavioral, and cognitive levels will not go unrewarded.

Notes

1. The development of chronometric measurement in the physical and then psychological sciences shows a parallel history. Bessel's 1820 discovery that astronomers differed systematically in determining the time of transit of a star led to two successive developments (Boring, 1950, p. 136): First, physical devices were perfected in order to reduce and then eliminate dependence on the "personal equation" of the human observer and, second, experimental psychologists, starting with Donders in 1868 and continuing to the present day (Sternberg, 1969) have turned such improved chronometric techniques back onto the problem of studying the temporal course of information processing within the human subject.

2. *Gedanken* experiments on psychophysical scaling with animals as subjects can be valuable in this connection. Such experiments confront us with the issue of how the experimenter is to train the subject to make responses that will somehow reveal the magnitudes of the subject's sensory experiences without at the same time biasing the subject toward some particular psychophysical relationship.

3. In the logarithmic model just as much as in the power model the parameter β must be regarded as an empirically determinable property of each continuum and not merely as an arbitrary scale factor. Thus, as Treisman has noted, the supposition by Stevens (1964) that the predictability of the interrelations among the exponents discovered in cross-modality matching would be sacrificed in the case of the logarithmic model appears unfounded. (Also see Luce & Galanter, 1963, p. 280.)

4. In revising the present section, I have adopted Krantz's now relatively entrenched term "relation theory" instead of speaking, as I did in my original 1968 draft, of "interpretation via a model for relative psychophysical judgment." Otherwise, I have retained my original exposition and derivations rather than attempting a reformulation along the

somewhat different, though in parts more elegant, formulation offered by Krantz (1972a). Krantz's Equations (18) and (12) correspond, respectively, to my invariance and transitivity Equations (6) and (13) and, from these, he arrives at essentially the same conclusions as I did. Also see Krantz (1972b, 1974) for his further work along related lines.

5. After this section was originally written, Ross and DiLollo (1968, 1971) independently reported new psychophysical results and analyses (including some theoretical developments related to those described here) that, however, complicate the unidimensionally consistent picture presented above. In the case of magnitude estimation of heaviness of lifted weights in particular, Ross and DiLollo found that the dimension that is effectively being judged tends to shift back and forth between the nonlinearly related attributes of weight and density depending upon the context of recently presented weights. As they noted, other continua might also be susceptible to such shifts in the basis for judgment (as, in the case of size judgments, between linear extent and area). In such cases, the above claim that all the judged ratios should be consistent with each other might need some modification.

6. In the original 1968 version of this paper, I had also explored the possibility that, in some tasks, subjects might equate differences rather than ratios of physical magnitudes. In preparing the present version, I have eliminated that discussion because I believe that the treatment of this possibility by Krantz, Luce, Suppes, and Tversky (1971) is much more satisfactory. The judgment of sense distance as opposed to sense ratios underlies widely used methods of multidimensional scaling. It is noteworthy that, whereas I conclude here that merely ordinal information about ratios does not permit the recovery of a ratio scale of magnitude, in the context of nonmetric multidimensional scaling, I have shown that merely ordinal information about distances does permit the recovery of a ratio scale of distance (Shepard, 1962, 1966).

7. After independently noting the possibility of this type of scale, I proposed to Stevens (personal communication, 1956) that, for overall consistency in his classificatory scheme, the designation "ratio scale" should properly be reserved for scales of this new type (in which equivalences are defined for ratios only) while scales that he had been designating in this way should more properly be called "interval-ratio scales" (since, in them, equivalences are defined for both differences and ratios). However, when he subsequently extended his classificatory scheme to encompass scales of this new variety (Stevens, 1957), he preferred to introduce, for them, the new term "logarithmic interval scales," in order to avoid a departure from the earlier usage of the term "ratio scale."

References

- Anderson, N. H. Functional measurement and psychophysical judgment. *Psychological Review*, 1970, 77, 153-170.
- Becker, R. *Theory of heat*. New York: Springer-Verlag, 1967.
- Birnbaum, M. H., & Veit, C. T. Scale convergence as a criterion for rescaling: Information integration with difference, ratio, and averaging tasks. *Perception & Psychophysics*, 1974, 15, 7-15.
- Boring, E. G. *A history of experimental psychology*. New York: Appleton-Century-Crofts, 1950.
- Carroll, J. D. *Functional learning: The learning of continuous functional mappings relating stimulus and response continua* (ETS RB 63-26). Princeton, N.J.: Educational Testing Service, 1963.
- Cohen, W. Spatial and textural characteristics of the Ganzfeld. *American Journal of Psychology*, 1957, 70, 403-410.

- Cornsweet, T. N. *Visual perception*. New York: Academic Press, 1970.
- Curtis, D. W., Paulos, M. A., & Rule, S. J. Relation between disjunctive reaction time and stimulus difference. *Journal of Experimental Psychology*, 1973, 99, 167-173.
- Ekman, G. Is the power law a special case of Fechner's law? *Perceptual and Motor Skills*, 1964, 19, 730.
- Falmagne, J.-C. The generalized Fechner problem and discrimination. *Journal of Mathematical Psychology*, 1971, 8, 22-43.
- Falmagne, J.-C. Random conjoint measurement and loudness summation. *Psychological Review*, 1976, 83, 65-79.
- Fechner, G. T. *Elemente der Psychophysik*. Leipzig: Breitkopf und Hartel, 1860. (Translation of Volume I reprinted as *Elements of psychophysics*, New York: Holt, Rinehart & Winston, 1966.)
- Feigl, H. The "mental" and the "physical." In H. Feigl & M. Scriven (Eds.), *Minnesota studies in the philosophy of science* (Vol. II). Minneapolis: University of Minnesota Press, 1958.
- Garner, W. R. An equidiscriminability scale for loudness judgments. *Journal of Experimental Psychology*, 1952, 43, 232-238.
- Garner, W. R. Context effects and the validity of loudness scales. *Journal of Experimental Psychology*, 1954, 48, 218-224.
- Graham, C. H. Sensation and perception in an objective psychology. *Psychological Review*, 1958, 65, 65-76.
- Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.
- Guttman, N., & Kalish, H. I. Discriminability and stimulus generalization. *Journal of Experimental Psychology*, 1956, 51, 79-88.
- Helson, H. *Adaptation level theory*. New York: Harper & Row, 1964.
- Herrnstein, R. J., & Boring, E. G. (Eds.). *A source book in the history of psychology*. Cambridge, Mass.: Harvard University Press, 1966.
- Hochberg, J. E., Treibel, W., & Seaman, G. Color adaptation under conditions of homogeneous visual stimulation (Ganzfeld). *Journal of Experimental Psychology*, 1951, 41, 153-159.
- Hull, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.
- Kiang, N. Y.-S. *Discharge patterns of single fibers in the cat's auditory nerve*. Cambridge, Mass.: M.I.T. Press, 1965.
- Klemmer, E. T., & Shrimpton, N. Preference scaling via a modification of Shepard's proximity analysis method. *Human Factors*, 1963, 5, 163-168.
- Krantz, D. H. A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology*, 1972, 9, 168-199. (a)
- Krantz, D. H. Measurement structures and psychological laws. *Science*, 1972, 175, 1427-1435. (b)
- Krantz, D. H. Measurement theory and qualitative laws in psychophysics. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology*. San Francisco: Freeman, 1974.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. *Foundations of measurement* (Vol. 1). New York: Academic Press, 1971.
- Kruskal, J. B., & Shepard, R. N. A nonmetric variety of linear factor analysis. *Psychometrika*, 1974, 39, 123-157.
- Land, E. H. The Retinex. *American Scientist*, 1964, 52, 247-264.

- Land, E. H. Color vision: From retina to Retinex. William James Lectures, presented at Harvard University, November 1966.
- Levelt, W. J. M., Riemersma, J. B., & Bunt, A. A. Binaural additivity of loudness. *British Journal of Mathematical and Statistical Psychology*, 1972, 25, 51-68.
- Levine, M. V. Transformations that render curves parallel. *Journal of Mathematical Psychology*, 1970, 7, 410-443.
- Levine, M. V. Transforming curves into curves with the same shape. *Journal of Mathematical Psychology*, 1972, 9, 1-16.
- Luce, R. D. *Individual choice behavior: A theoretical analysis*. New York: Wiley, 1959.
- Luce, R. D. What sort of measurement is psychophysical measurement? *American Psychologist*, 1972, 27, 96-106.
- Luce, R. D., & Galanter, E. Psychophysical scaling. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1). New York: Wiley, 1963.
- 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. & Tukey, J. Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1964, 1, 1-27.
- MacKay, D. M. Psychophysics of perceived intensity: A theoretical basis for Fechner's and Stevens' laws. *Science*, 1963, 139, 1213-1216.
- McCann, J. J. Rod-cone interactions: Different color sensations from identical stimuli. *Science*, 1972, 176, 1255-1257.
- McGill, W. The slope of the loudness function: A puzzle. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: Theory and applications*. New York: Wiley, 1960.
- Middleton, W. E. K. *A history of the thermometer and its uses in meteorology*. Baltimore: Md.: Johns Hopkins Press, 1966.
- Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, 63, 81-97.
- Mountcastle, V. B., Poggio, G. F., & Werner, G. The relation of thalamic cell response to peripheral stimuli varied over an intensive continuum. *Journal of Neurophysiology*, 1963, 26, 807-834.
- Oyamo, T. A behavioristic analysis of Stevens' magnitude estimation method. *Perception & Psychophysics*, 1968, 3, 317-320.
- Pattie, F. A. A report of attempts to produce uniocular color blindness by hypnotic suggestion. *British Journal of Medical Psychology*, 1935, 15, 230-241.
- Perkel, D. H. & Bullock, T. H. Neural coding. *Bulletin, Neurosciences Research Program*, 1968, 6, 221-348.
- Pickersgill, M. J. On knowing with which eye one is seeing. *Quarterly Journal of Experimental Psychology*, 1961, 13, 168-172.
- Pollack, I. Iterative techniques for unbiased rating scales. *Quarterly Journal of Experimental Psychology*, 1965, 17, 139-148.
- Powers, W. T. *Behavior: The control of perception*. Chicago: Aldine, 1973.
- Rock, I. *An introduction to perception*. New York: Macmillan, 1975.
- Rock, I., & Ebenholtz, S. The relational determination of perceived size. *Psychological Review*, 1959, 66, 387-401.
- Ross, J., & DiLollo, V. A vector model for psychophysical judgment. *Journal of Experimental Psychology Monograph Supplement*, 1968, 77, (3, Part 2).
- Ross, J., & DiLollo, V. Judgment and response in magnitude estimation. *Psychological Review*, 1971, 78, 515-527

- Rushton, W. A. H. Peripheral coding in the nervous system. In W. A. Rosenblith (Ed.), *Sensory communication*. Cambridge, Mass.: M.I.T. Press, 1961.
- Savage, C. W. *The measurement of sensation*. Berkeley: University of California Press, 1970.
- Selfridge, O. G. Pandemonium, a paradigm for learning. In *Proceedings of symposium on the mechanization of thought processes*. National Physical Laboratory, Teddington, England. London: H. M. Stationery Office, 1958.
- Shepard, R. N. Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychological Review*, 1958, 65, 242-256. (a)
- Shepard, R. N. Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology*, 1958, 55, 509-523. (b)
- Shepard, R. N. Role of generalization in stimulus-response compatibility. *Perceptual and Motor Skills*, 1961, 13, 59-62.
- Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. I & II. *Psychometrika*, 1962, 27, 125-140, 219-246.
- Shepard, R. N. Computers and thought: A review of the book edited by Feigenbaum and Feldman. *Behavioral Science*, 1964, 9, 57-65.
- Shepard, R. N. Approximation to uniform gradients of generalization by monotone transformations of scale. In D. I. Mostofsky (Ed.), *Stimulus generalization*. Stanford, Calif.: Stanford University Press, 1965.
- Shepard, R. N. Metric structures in ordinal data. *Journal of Mathematical Psychology*, 1966, 3, 287-315.
- Shepard, R. N. Form, formation, and transformation of internal representations. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975.
- Shepard, R. N., & Carroll, J. D. Parametric representation of nonlinear data structures. In P. R. Krishnaiah (Ed.), *Multivariate analysis*. New York: Academic Press, 1966.
- Shepard, R. N., & Chang, J.-J. Stimulus generalization in the learning of classifications. *Journal of Experimental Psychology*, 1963, 65, 94-102.
- Shepard, R. N., Kilpatrick, D. W., & Cunningham, J. P. The internal representation of numbers. *Cognitive Psychology*, 1975, 7, 82-138.
- Siegel, J. A. The nature of absolute pitch. In I. Gordon (Ed.), *Studies in the psychology of music* (Vol. 8). Iowa City: University of Iowa Press, 1972.
- Skinner, B. F. The operational analysis of psychological terms. *Psychological Review*, 1945, 52, 270-277.
- Skinner, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.
- Sternberg, S. Four theories of the effect of distance on apparent size. Paper read at the Psychonomic Society meeting, St. Louis, October, 1966.
- Sternberg, S. The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 1969, 30, 276-315.
- Stevens, J. C. & Stevens, S. S. Warmth and cold—dynamics of sensory intensity. *Journal of Experimental Psychology*, 1960, 60, 183-192.
- 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.
- Stevens, S. S. The direct estimation of sensory magnitudes—Loudness. *American Journal of Psychology*, 1956, 69, 1-25.
- 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. Ratio scales, partition scales, and confusion scales. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: Theory and applications*. New York: Wiley, 1960.
- Stevens, S. S. To honor Fechner and repeal his law. *Science*, 1961, 133, 80-86.
- Stevens, S. S. Concerning the psychophysical power law. *Quarterly Journal of Experimental Psychology*, 1964, 16, 383-385.
- Stevens, S. S. Matching functions between loudness and ten other continua. *Perception & Psychophysics*, 1966, 1, 5-8. (a)
- Stevens, S. S. On the operation known as judgment. *American Scientist*, 1966, 54, 385-401. (b)
- Stevens, S. S., & Guirao, M. Subjective scaling of length and area and the matching of length to loudness and brightness. *Journal of Experimental Psychology*, 1963, 66, 177-186.
- Stevens, S. S., & Guirao, M. The scaling of apparent viscosity. *Science*, 1964, 144, 1157-1158.
- Suppes, P., & Zinnes, J. L. Basic measurement theory. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. I). New York: Wiley, 1963.
- Teghtsoonian, M. The judgment of size. *American Journal of Psychology*, 1965, 78, 392-402.
- Thouless, R. H. Phenomenal regression to the "real" object. I & II. *British Journal of Psychology*, 1931, 21, 339-359; 22, 1-30.
- Thurstone, L. L. A law of comparative judgment. *Psychological Review*, 1927, 34, 273-286.
- Torgerson, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.
- Torgerson, W. S. Quantitative judgment scales. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: Theory and applications*. New York: Wiley, 1960.
- Treisman, M. Laws of sensory magnitude. *Nature*, 1963, 198, 914-915.
- Treisman, M. Sensory scaling and the psychophysical law. *Quarterly Journal of Experimental Psychology*, 1964, 16, 11-22.
- Ward, W. D. Absolute pitch. *Sound*, 1963, Part I, 2, 14-41; Part II, 3, 33-41.
- Warren, R. W. A basis for judgment of sensory intensity. *American Journal of Psychology*, 1958, 71, 675-687.
- Warren, R. M., Sersen, E. A., & Pores, E. B. A basis for loudness judgments. *American Journal of Psychology*, 1958, 71, 700-709.
- Welford, A. T. The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics*, 1960, 3, 189-230.
- Wickelgren, W. A. Associative strength theory of recognition memory for pitch. *Journal of Mathematical Psychology*, 1969, 6, 13-61.
- Wittgenstein, L. *Philosophical investigations* (G. E. M. Anscombe, trans.). New York: Macmillan, 1953.
- Woodworth, R. S. *Experimental psychology*. New York: Henry Holt, 1938.