Psychophysical Scaling

Hans Irtel

Volume 3, pp. 1628–1632

in

Encyclopedia of Statistics in Behavioral Science

ISBN-13: 978-0-470-86080-9
ISBN-10: 0-470-86080-4

Editors

Brian S. Everitt & David C. Howell

Psychophysical Scaling

Introduction

A large part of human cognition is devoted to the development of a mental representation of the physical environment. This is necessary for planned interaction and for successful anticipation of dangerous situations. Survival in a continuously changing physical world will only be possible if the organism's mental representation of its environment is sufficiently valid. This actually is the case for most of our perceptual abilities. Our senses convey a rather valid view of the physical world, at least as long as we restrict the world to that part of the environment that allows for unaided interaction. And this is the basic idea of psychophysical scaling: Formulate a theory that allows the computation of perceived stimulus properties from purely physical attributes.

A problem complex as this requires simplification and reduction in order to create proper experiments and models that can be used to describe, or even explain, the results of these experiments. Thus, most of the experimental examples and most of the theories we will discuss here will be small and simplified cases. Most cases will assume that we have only a single relevant physical independent variable, and a single and unidimensional dependent variable. Simplification like this is appropriate as long as we keep in mind the bigger aim of relating the properties of the physical world to behavior, or, better, to the mental representation of the world that guides behavior.

Subject Tasks

Before going into the theoretical foundations of psychophysical scaling it might be useful to look at the experimental conditions which give rise to problems of psychophysical scaling. So we first will look at subject tasks which may be found in experiments involving psychophysical scaling. The most basic tasks involved in psychophysical experiments are detection and discrimination tasks. A *detection* task involves one or more time intervals during which a stimulus may be presented. The subject's task is to tell which if any time interval contained the target stimulus. Single time interval tasks usually are called *yes/no*-tasks while multiple time interval tasks are called *forced choice*-tasks. *Discrimination* tasks are similar to multiple interval detection tasks. The major difference being that there are no empty intervals but the distractor intervals contain a reference or standard stimulus and the subject's task is to tell, whether the target is different from the reference or which of multiple intervals contains the target. The data of detection and discrimination experiments usually are captured by the probability that the respective target is detected or discriminated from its reference.

When scaling is involved, then discrimination frequently involves ordering. In this case, the subject is asked whether the target stimulus has more of some attribute as a second target that may be presented simultaneously or subsequently. These tasks are called *paired comparison* (*see* **Bradley–Terry Model**) tasks, and frequently involve the comparison of stimulus pairs. An example may be a task where the difference between two stimuli x, y with respect to some attribute has to be compared with the difference between two stimuli u, v. The data of these comparisons may also be handled as probabilities for finding a given ordering between pairs.

Another type of psychophysical task involves the assignment of numeric labels to stimuli. A simple case is the assignment of single stimuli to a small number of numeric categories. Or subjects may be required to directly assign real number labels to stimuli or pairs of stimuli such that the numbers describe some attribute intensity associated with physical stimulus properties. The most well-known task of this type is *magnitude estimation*: here the subject may be asked to assign a real number label to a stimulus such that the real number describes the appearance of some stimulus attribute such as loudness, brightness, or heaviness. These tasks usually involve stimuli that show large differences with respect to the independent physical attribute such that discrimination would be certain if two of them were presented simultaneously. In many cases, the numerical labels created by the subjects are treated as numbers such that the data will be mean values of subject responses.

The previously described task type may be modified such that the subject does not produce a number label but creates a stimulus attribute such that its appearance satisfies a certain numeric relation to a given reference. An example is *midpoint production*: here the subject adjusts a stimulus attribute such that it has an intensity that appears to be the midpoint between two given reference stimuli. Methods of this type are called *production methods* since the subject produces the respective stimulus attribute. This is in contrast to the *estimation methods*, where the subjects produce a numerical estimation of the respective attribute intensity.

Discrimination Scaling

Discrimination scaling is based on an assumption that dates back to **Fechner** [3]. His idea was that psychological measurement should be based on discriminability of stimuli. Luce & Galanter [5] used the phrase 'equally often noticed differences are equal, unless always or never noticed' to describe what they called *Fechner's Problem*:

Suppose $P(x, y)$ is the discriminability of stimuli x and y. Does there exist a transformation g of the physical stimulus intensities x and y, such that

$$P(x, y) = F[g(x) - g(y)], \tag{1}$$

where F is a strictly increasing function of its argument? Usually, $P(x, y)$ will be the probability that stimulus x is judged to be of higher intensity as stimulus y with respect to some attributes. A solution g to (1) can be considered a psychophysical scale of the respective attribute in the sense that equal differences along the scale g indicate equal discriminability of the respective stimuli. An alternative but empirically equivalent formulation of (1) is

$$P(x, y) = G\left[\frac{h(x)}{h(y)}\right], \tag{2}$$

where $h(x) = e^{g(x)}$ and $G(x) = F(\log x)$.

Response probabilities $P(x, y)$ that have a representation like (1) have to satisfy the *quadruple condition*: $P(x, y) \geq P(u, v)$ if and only if $P(x, u) \geq P(y, v)$. This condition, however, is not easy to test since no statistical methods exist that allow for appropriate decisions based on estimates of P.

Response probabilities that have a Fechnerian representation in the sense of (1) allow the definition of a *sensitivity function* ξ: Let $P(x, y) = \pi$ and define $\xi_\pi(y) = x$. Thus, $\xi_\pi(y)$ is that stimulus intensity which, when compared to y, results in response probability π. From (1) with $h = F^{-1}$, we get

$$\xi_\pi(y) = g^{-1}[g(y) + h(\pi)], \tag{3}$$

where $h(\pi)$ is independent of x.

Fechner's idea was that a fixed response probability value π corresponds to a single unit change on the sensation scale g: We take $h(\pi) = 1$ and look at the so called *Weber function* δ, which is such that $\xi_\pi(x) = x + \delta_\pi(x)$. The value of the Weber function $\delta_\pi(x)$ is that stimulus increment that has to be added to stimulus x such that the response probability $P(x + \delta_\pi(x), x)$ is π. *Weber's law* states that $\delta_\pi(x)$ is proportional to x [16]. In terms of response probabilities, this means that $P(cx, cy) = P(x, y)$ for any multiplicative factor c. Weber's law in terms of the sensitivity function ξ means that $\xi_\pi(cx) = c\xi_\pi(x)$. A generalization is $\xi_\pi(cx) = c^\beta \xi_\pi(x)$, which has been termed the *near-miss-to-Weber's-law* [2]. Its empirical equivalent is $P(c^\beta x, cy) = P(x, y)$ with the representation $P(x, y) = G(y/x^{1/\beta})$. The corresponding Fechnerian representation then has the form

$$P(x, y) = F\left[\frac{1}{\beta}\log x - \log y\right]. \tag{4}$$

In case of $\beta = 1$, we get *Fechner's law* according to which the sensation scale grows with the logarithmic transformation of stimulus intensity.

Operations on the Stimulus Set

Discrimination scaling is based on stimulus confusion. Metric information is derived from data that somehow describe a subject's uncertainty when discriminating two physically distinct stimuli. There is no guarantee that the concatenation of just noticeable differences leads to a scale that also describes judgments about stimulus similarity for stimuli that are never confused. This has been criticized by [14].

An alternative method for scale construction is to create an operation on the set of stimuli such that this operation provides metric information about the appearance of stimulus differences. A simple case is the midpoint operation: the subject's task is to find that stimulus $m(x, y)$, whose intensity appears to be the midpoint between the two stimuli x and y. For any sensation scale g, this should mean that

$$g[m(x, y)] = \frac{g(x) + g(y)}{2}. \tag{5}$$

The major empirical condition that guarantees that the midpoint operation m may be represented in this way is *bisymmetry* [11]:

$$m[m(x, y), m(u, v)] = m[m(x, u), m(y, v)], \tag{6}$$

which can be tested empirically. If bisymmetry holds, then a representation of the form

$$g[m(x, y)] = pg(x) + qg(y) + r \qquad (7)$$

is possible. If, furthermore, $m(x, x) = x$ holds, then $p + q = 1$ and $r = 0$, and if $m(x, y) = m(y, x)$, then $p = q = 1/2$. Note that bisymmetry alone does not impose any restriction on the form of the psychophysical function g. However, Krantz [4] has shown that an additional empirically testable condition restricts the possible forms of the psychophysical function strongly. If the sensation scale satisfies (5), and the midpoint operation satisfies the homogeneity condition $m(cx, cy) = cm(x, y)$, then there remain only two possible forms of the psychophysical function g: $g(x) = \alpha \log x + \beta$, or $g(x) = \alpha x^{\beta} + \gamma$, for two constants $\alpha > 0$ and β. Falmagne [2] makes clear that the representation of m by an arithmetic mean is arbitrary. A geometric mean would be equally plausible, and the empirical conditions given above do not allow any distinction between these two options. Choosing the geometric mean as a representation for the midpoint operation m, however, changes the possible forms of the psychophysical function g [2].

The midpoint operation aims at the single numerical weight of $1/2$ for multiplying sensation scale values. More general cases have been studied by [9] in the context of magnitude estimation, which will be treated later. Methods similar to the midpoint operation have also been suggested by [14] under the label *ratio* or *magnitude production* with *fractionation* and *multiplication* as subcases. Pfanzagl [11], however, notes that these methods impose much less empirical constraints on the data such that almost any monotone psychophysical function will be admissible.

Magnitude Estimation, Magnitude Production, and Cross-modal Matching

Magnitude estimation is one of the classical methods proposed by [14] in order to create psychophysical scales that satisfy proper measurement conditions. Magnitude estimation requires the subject to assign numeric labels to stimuli such that the respective numbers are proportional to the magnitude of perceived stimulus intensity. Often, there will be a reference stimulus that is assigned a number label, '10', say, by the experimenter, and the subject is instructed to map the relative sensation of the target with respect to the reference. It is common practice to take the subjects' number labels as proper numbers and compute average values from different subjects or from multiple replications with the same subject. This procedure remains questionable as long as no structural conditions are tested that validate the subjects' proper handling of numeric labels. In magnitude production experiments, the subject is not required to produce number labels but to produce a stimulus intensity that satisfies a given relation on the sensation continuum to a reference, such as being 10 times as loud.

A major result of **Stevens'** research tradition is that average data from magnitude estimation and production frequently are well described by power functions: $g(x) = \alpha x^{\beta}$. Validation of the power law is done by fitting the respective power function to a set of data, and by *cross-modal matching*. This requires the subject to match a pair of stimuli from one continuum to a second pair of stimuli from another continuum. An example is the matching of a loudness interval defined by a pair of acoustic stimuli to the brightness interval of a pair of light stimuli [15]. If magnitude estimates of each single sensation scale are available and follow the power law, then the exponent of the matching function from one to the other continuum can be predicted by the ratio of the exponents. Empirical evidence for this condition is not unambiguous [2, 7]. In addition, a power law relation between matching sensation scales for different continua is also predicted by logarithmic sensation scales for the single continua [5].

The power law for sensation scales satisfies Stevens' credo 'that equal stimulus ratios produce equal subjective ratios' ([14], p 153). It may, in fact, be shown that this rule implies a power law for the sensation scale g. But, as shown by several authors [2, 5], the notion 'equal subjective ratios' has the same theoretical status as Fechner's assumption that just noticeable stimulus differences correspond to a single unit difference on the sensation scale. Both assumptions are arbitrary as long as there is no independent foundation of the sensation scale.

A theoretical foundation of magnitude estimation and cross-modal matching has been developed by [4] based on ideas of [12]. He combined magnitude estimates, ratio estimates, and cross-modal matches, and formulated a set of empirically testable conditions that have to hold if the psychophysical functions are power functions of the corresponding physical attribute. A key assumption of the Shepard–Krantz theory is to map all sensory attributes to the single sensory continuum of perceived length, which is assumed to behave like physical length. The main assumption, then, is that for the reference continuum of perceived length, the condition $L(cx, cy) = L(x, y)$ holds. Since, furthermore, $L(x, y)$ is assumed to behave like numerical ratios, this form of invariance generates the power law [2].

A more recent theory of magnitude estimation for ratios has been developed by [9]. The gist of this theory is a strict separation of number labels, as they are used by subjects, and mathematical numbers used in the theory. Narens derived two major predictions: The first is a commutativity property. Let 'p' and 'q' be number labels, and suppose the subject produces stimulus y when instructed to produce p-times x, and then produces z when instructed to produce q-times y. The requirement is that the result is the same when the sequence of 'p' and 'q' is reversed. The second prediction is a multiplicative one. It requires that the subject has to produce the same stimulus as in the previous sequence when required to produce pq-times x. Empirical predictions like this are rarely tested. Exceptions are [1] or for a slightly different but similar approach [18]. In both cases, the model was only partially supported by the data.

While [6] describes the foundational aspects of psychophysical measurement, [8] presents a thorough overview over the current practices of psychophysical scaling in Stevens' tradition. A modern characterization of Stevens' measurement theory [13] is given by [10].

For many applications, the power law and the classical psychophysical scaling methods provide a good starting point for the question how stimulus intensity transforms into sensation magnitude. Even models that try to predict sensation magnitudes in rather complex conditions may incorporate the power law as a basic component for constant context conditions. An example is the CIE 1976 ($L^*a^*b^*$) color space [17]. It models color similarity judgments and implements both an adaptation- and an illumination-dependent component. Its basic transform from stimulus to sensation space, however, is a power law.

References

[1] Ellermeier, W. & Faulhammer, G. (2000). Empirical evaluation of axioms fundamental to Stevens' ratio-scaling approach: I. loudness production, *Perception & Psychophysics* **62**, 1505–1511.

[2] Falmagne, J.-C. (1985). *Elements of Psychophysical Theory*, Oxford University Press, New York.

[3] Fechner, G.Th. (1860). *Elemente der Psychophysik*, Breit-kopf und Härtel, Leipzig.

[4] Krantz, D.H. (1972). A theory of magnitude estimation and cross-modality matching, *Journal of Mathematical Psychology* **9**, 168–199.

[5] Luce, R.D. & Galanter, E. (1963). Discrimination, in *Handbook of Mathematical Psychology*, Vol. I, R.D. Luce, R.R. Bush & E. Galanter, eds, Wiley, New York, pp. 191–243.

[6] Luce, R.D. & Suppes, P. (2002). Representational measurement theory, in *Stevens' Handbook of Experimental Psychology*, Vol. 4, 3rd Edition, H. Pashler, ed., Wiley, New York, pp. 1–41.

[7] Marks, L.E. & Algom, D. (1998). Psychophysical scaling, in *Measurement, Judgement, and Decision Making*, M.H. Birnbaum, ed., Academic Press, San Diego, pp. 81–178.

[8] Marks, L.E. & Gescheider, G.A. (2002). Psychophysical scaling, in *Stevens' Handbook of Experimental Psychology*, H. Pashler, ed., Vol. 4, 3rd Edition, Wiley, New York, pp. 91–138.

[9] Narens, L. (1996). A theory of ratio magnitude estimation, *Journal of Mathematical Psychology* **40**, 109–129.

[10] Narens, L. (2002). The irony of measurement by subjective estimations, *Journal of Mathematical Psychology* **46**, 769–788.

[11] Pfanzagl, J. (1971). *Theory of Measurement*, 2nd Edition, Physica-Verlag, Würzburg.

[12] Shepard, R.N. (1981). Psychophysical relations and psychophysical scales: on the status of 'direct' psychophysical measurement, *Journal of Mathematical Psychology* **24**, 21–57.

[13] Stevens, S.S. (1946). On the theory of scales of measurement, *Science* **103**, 677–680.

[14] Stevens, S.S. (1957). On the psychophysical law, *The Psychological Review* **64**, 153–181.

[15] Stevens, J.C. & Marks L.E. (1965). Cross-modality matching of brightness and loudness, *Proceedings of the National Academy of Sciences*, **54**, 407–411.

[16] Weber, E.H. (1834). *De Pulsu, Resorptione, Auditu et Tactu. Annotationes Anatomicae et Physiologicae*, Köhler, Leipzig.

[17] Wyszecki, G. & Stiles, W.S. (1982). *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd Edition, Wiley, New York.

[18] Zimmer, K., Luce, R.D. & Ellermeier, W. (2001). Testing a new theory of psychophysical scaling: temporal loudness integration, in *Fechner Day 2001. Proceedings of the Seventeenth Annual Meeting of the International Society for Psychophysics*, E. Sommerfeld, R. Kompass & T. Lachmann, eds, Pabst, Lengerich, pp. 683–688.

(*See also* **Harmonic Mean**)

HANS IRTEL