MASIN, S. C. (2012). A brief trip into the history of psychophysical measurement. In C. Leth-Steensen, & J. R. Shoenherr (Eds.), *Proceedings of the twenty-eighth annual meeting of the International Society for Psychophysics*. Ottawa, ON: International Society for Psychophysics (pp. 162-167).

A BRIEF TRIP INTO THE HISTORY OF PSYCHOPHYSICAL MEASUREMENT

Sergio Cesare Masin
University of Padua, Italy
scm@unipd.it

Abstract

Psychophysical measurement was first used for scientific purposes more than 2000 years ago. Its development is overviewed to the present days with emphasis on the most promising lead.

Psychophysical measurement was used for scientific purposes for the first time by Hipparchus of Rhodes around 150 BCE. We reach this conclusion as follows.

In his encyclopedia *Naturalis Historia* (about 80 CE) Pliny the Elder wrote the following: “Hipparchus the foresaid Philosopher (a man never sufficiently praised, as who proved the affinitie of starres with men, and none more than he, affirming also, that our soules were parcell of heaven) found out and observed another new starre engendred in his time, and by the motion thereof on what day it first shone, he grew presently into a doubt, Whether it happened not very often that new starres should arise? and whether those starres also mooved not, which we imagine to be fast fixed? The same man went so farre, that he attempted (a thing even hard for God to performe) to deliver unto posteritie the just number of starres. Hee brought the said starres within the compasse of rule and art, devising certaine instruments to take their severall places, and set out their magnitudes: that thereby it might be easily discerned, not only whether the old died, and new were borne, but also whether they moved, and which way they tooke their course? likewise, whether they encreased or decreased? Thus he left the inheritance of heaven unto all men, if any one haply could be found able to enter upon it as lawfull heire” (Book II, Ch. 26, in Holland, 1601).

Thus Hipparchus recorded positions and magnitudes of stars to allow posterity to determine not only whether old stars had died and new stars were born but also whether stars had moved and whether their magnitude had changed. Pliny the Elder says that Hipparchus measured star position by instruments. He does not say how Hipparchus measured stellar magnitude. We can obtain the missing information from Ptolemy’s star catalogue published in the *Almagest* (about 150 CE) since this catalogue used most probably earlier work by Hipparchus that is now lost. The catalogue gives coordinates and magnitudes of many stars. Stellar magnitude is “the class” (Book 7, Ch. 4, in Toomer, 1984) to which a star belongs in terms of perceived brightness. As we know, stellar magnitudes vary from I (brightest) to VI (dimmest).

Hipparchus could only measure perceived brightness since he had no photometer. For this he resolved to use category rating, apparently the most natural method of mental measurement. Variants of this method were subsequently used to measure magnitudes of stars in the telescope (Hearnshaw, 1996).

In 1740, Celsius and Tulenius were the first to obtain photometric measures of relative starlight intensity (Weaver, 1946) but it was only in the early 1800’s that John Herschel (1829, p. 182) and Steinheil (1837) could provide sufficiently accurate measures. Their measurements showed for the first time that the relation between rated stellar brightness and physical relative starlight intensity was approximately logarithmic (Hearnshaw, 1996, p. 76).

In 1840, Plateau measured perceived lightness using his well-known method of bisection (Plateau, 1872). The method consists in defining an initial sensory interval delimited by two largely different perceived magnitudes of a sensation and by having a person produce equidifferent perceived magnitudes that divide the initial interval in equal subintervals. Each of these subintervals is taken as the mental unit of measurement.

Fechner (1860) proposed the following method for measuring those perceived sensations that co-vary with a known physical variable. For each fixed value of the physical variable, S, one determines the increment Δs of S that produces the smallest possible difference in the sensation. The smallest possible difference in sensation is assumed to be invariant with S. One determines the best-fitting function relating Δs to S, called the Weber function. With S_0 denoting the minimum S that evokes a sensation, one adds Δs to S_0, to the S resulting from this first addition, to the S resulting from this second addition, and so on, each time using the Weber function to select the Δs to be added to a new S. The number of additions of Δs to S_0 necessary to reach the S that produces the perceived magnitude being measured is the measure of this magnitude. Each addition defines one mental unit of measurement.

In 1887, Fechner argued that the rating method, the bisection method, and his own method produce acceptable measures of perceived sensation. He also argued that his own method should be preferable since it produces ratio-scale measures while the rating and bisection methods produce interval-scale measures (Scheerer, 1987).

Merkel (1888) and Fullerton and Cattell (1892) proposed the method of measurement in which a person selects a variable stimulus such that its perceived magnitude is in a fixed ratio with that of a standard stimulus. The perceived magnitude of the standard defines the mental unit of measurement. The bisection and Merkel's methods assume people's ability to judge the equalities of differences and of ratios of perceived magnitudes, respectively.

In 1921, Brown and Thomson set forth the central idea of the method of measurement today called nonmetric scaling: "To take a simple example, suppose five quantities *a*, *b*, *c*, *d*, *e* have really the measures 10, 16, 20, 31, 32." Have a person ignorant of these measures rank first differences $|a - b|$, $|b - c|$, $|c - d|$, ... , second differences $|d - c| - |b - a|$, $|c - b| - |e - d|$, $|b - a| - |c - b|$ ... , or even third differences. "If now we could have all these [rankings] we could space out the original quantities very closely indeed to their true positions. This can be best seen by attempting to alter some one of the values while leaving all [rankings] unaltered. Make *d*, for example, 29 instead of 31 and although the order *a*, *b*, *c*, *d*, *e* is unchanged, and also the order of the first differences, that of the second differences is completely altered (Brown & Thomson, 1921, p. 12)." The method is prohibitive since it requires a large number of stimulus trials even using only first differences (Shepard, 1966).

The methods of bisection, of Fechner, and of Merkel can only apply to mental variables that co-vary with a known physical variable. The usefulness of these methods is thus very limited. On the other hand, the rating method applies to any mental variable.

In 1929, these facts must have prompted Richardson to propose direct numerical magnitude estimation (Richardson, 1929a; Richardson & Ross, 1930) and graphic rating (Richardson, 1929b) to measure any mental magnitude. He measured strength of imagery by magnitude estimation, and saturation of red by graphic rating. Before, De Marchi (1925) used magnitude estimation to measure visual dot density. Magnitude estimation assumes that people can judge sensory ratios. Since the 1930s, Richardson's methods and variants thereof have been widely used up to today (Gescheider, 1997; Marks & Algom, 1998; Stevens, 1975).

The validity of the above methods depends on the truth at least of the most relevant assumptions on which they are based. The problem is that this truth is hard to ascertain. Most relevant assumptions are the abilities to equalize perceived differences in rating, bisection, Fechner's method, and nonmetric scaling and to judge sensory ratios in Merkel's and Richardson's methods and variants thereof. It is believed that one can test these assumptions by first axiomatically formalizing the operations that underlie the methods and then use these formalizations to draw empirically testable logico-mathematical consequences (Luce, 1972, 2002). Examples are the axiomatic formalizations of bisection (Pfanzagl, 1959), Fechner's method (Falmagne, 1985), and ratio judgment (Narens, 1996) among many others.

Unfortunately, tests derived from axiomatic formalizations involve serious difficulties. (*i*) Doubts about the truth of assumptions are transferred to the logico-mathematical consequences of the formalization. That is, the conclusion is reached that a method is valid or invalid based only on the trust one is willing to put in the correctness of the formalization. For example, Pfanzagl (1959) gave an axiomatic formalization of the bisection operation yielding the logico-mathematical consequence called bisymmetry condition. For about 50 years it has been given for granted that empirically testing the bisymmetry condition was fundamental to establish the validity of the method (Falmagne, 1974; Luce & Galanter, 1963). Instead, this condition is totally irrelevant for the purpose of testing this validity: it applies indifferently to any relative magnitude a person arbitrarily chooses to divide an interval (Masin & Toffalini, 2009). (*ii*) Tests of axioms derived from axiomatic formalizations are ordinal tests. They suffer from order effect. For example, given the sensory intensities a, b, and c, Fagot and Stewart (1969) had persons judge the ratios $R_{ab} = a / b$, $R_{bc} = b / c$, and $R_{ac} = a / c$. Consistent ratio judgments imply $R_{ac} = R_{ab} \cdot R_{bc}$. It turned out that $R_{ac} \neq R_{ab} \cdot R_{bc}$. This inequality could depend on order effects rather than revealing inability to judge sensory ratios. Control of order effects is inherently flawed since we ignore how the size of effects varies with stimulus intensity and with presentation order. (*iii*) Various other arguments conclude that the significance of axiomatic formalizations in psychology is virtually nil (Anderson, 1981, pp. 349–353; Cliff, 1992; Estes 1975; Schönemann, 1994)

How can one then determine the validity of a method of psychophysical measurement given that all available evidence indicates that tests based on axioms are insufficient? One answer comes from functional measurement theory (Anderson, 1982, pp. 246–251).

Our scientific knowledge about nature accrues through a continuous process of formulating tentative theories and critically testing them by selected methods of measurement. Although theories may survive critical tests, they remain theories and can only asymptotically be established as true by this converging corroboration (Popper, 1963). Since a law is a theory about a mathematical relation between variables and is tested by measurement methods that one selects among various other possible methods, validating the law and selecting its method of measurement are two related aspects of the same process of corroboration (Ellis, 1966). A method of measurement is tentatively valid when it yields measures for the variables involved in a (tentatively valid) law that are in the same mathematical relation as that which defines the law. The method is increasingly corroborated as it progressively agrees with other newly discovered laws. The following is an example of this corroboration process, analogous to the one in Anderson & Cuneo (1978). For other validation tests see Anderson (1996, pp. 94–96).

A large body of data from many judgment tasks indicates that people integrate information using mental operations such as adding, multiplying, averaging, weighted averaging, etc. (Anderson, 1981, 1996). It is theorized that since we are evolutionarily adapted to the everyday empirical world we integrate information about empirical physical phenomena using mental operations for information integration that match the mathematical relations between the variables involved in the respective empirical physical laws, as if we intuitively knew these laws (Anderson, 1983). This theory and its measurement methods are simultaneously validated by a corroboration process whose initial steps are as those exemplified next.

Consider a flat object on a horizontal board with object and board covered with sandpaper. The minimum force necessary to slide the object on the board is proportional to the sum of the grit numbers of the sandpapers of object and board. For each factorial combination of these grit numbers, and with object and board kept separate, Corneli and Vicovaro (2007) had persons rate the imagined friction of the object on the board after each person had felt with their fingers how coarse the surfaces of object and board were. Figure 1a shows mean

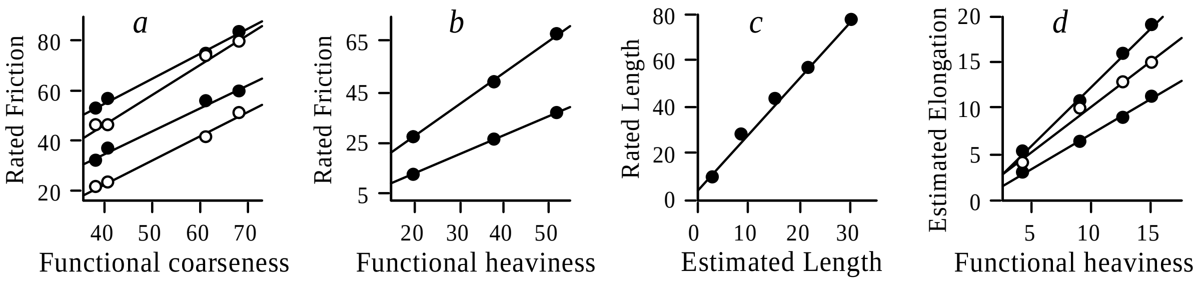

Fig. 1. Mean rated imagined friction against mean object functional coarseness (*a*) and functional heaviness (*b*), parameter: board coarseness (Corneli & Vicovaro, 2007); (*c*) mean rated length against mean estimated length (Masin, 2008); (*d*) mean estimated imagined elongation against mean object functional heaviness, parameter: spring length (Cocco & Masin, 2010).

rated friction against mean object functional coarseness for each board coarseness. [The mean of the mean ratings corresponding to each object coarseness is a mean functional measure of that felt object coarseness (Anderson, 1982, p. 58)]. The results agree with physical law.

These results tentatively support the theory that people implicitly know the additive physical law of friction and simultaneously validate the rating method used to measure imagined friction. The results suggest that ratings are linear measures of imagined friction.

These conclusions are tentative since they rest on one step only of the converging corroboration process. Some have misinterpreted that this process ends with this first step. For example, Gigerenzer and Richter (1990) argued that the same results may obtain if people multiply felt object coarseness by felt board coarseness and if ratings are logarithmic rather than linear measures of imagined friction. The next step overcomes this misinterpretation.

The minimum physical force needed to slide the object on the board equals the product of object weight by the friction coefficient. For each factorial combination of object weight and board coarseness, and with object and board kept separate, Corneli and Vicovaro (2007) had persons rate the imagined friction of the object on the board after they hefted the object and felt how coarse the board surface was. Figure 1b shows mean rated friction plotted against mean object functional heaviness for each board. The results agree with physical law.

These results further support the validity of the aforesaid theory and its measurement method—both tests involved the same method and the same measured variable but a different cognitive law. They reconfirm that ratings are linear measures: had ratings been logarithmic measures, factorial curves would have been parallel rather than being divergent.

Methods that yield linear measures are equivalent. This equivalence may hold only for some tasks. For example, ratings of average lightness are linear and nonequivalent to magnitude estimates (Weiss, 1972). On the other hand, for lengths in the range 2–68 cm, the results in Figure 1c show that ratings and magnitude estimates of apparent length are equivalent measures (Masin, 2008). Length estimation can thus be used to validate the above theory.

For a spring of length L hanging vertically, a load of weight W suspended from its lower end causes spring length to increase from L to $L + E$. The elongation E is proportional to the product $L \cdot W$. For different factorial combinations of L and W, Cocco and Masin (2010) had persons lift a load with their hands and, simultaneously, look at a spring and rate the imagined elongation of the spring that would occur in case the load was suspended on the lower end of the spring. Figure 1d shows the factorial curves relating mean estimated elongation to mean load functional heaviness for different Ls. The results agree with physical law.

These results reconfirm the validity of the aforesaid theory of implicit physical knowledge and its measurement method. They also further reconfirm that ratings and magnitude estimates of length are linear measures. This process of corroboration continues.

Convergent corroboration appears to be the only viable process of validation of psychophysical measurement. It is consequently desirable that more attempts are made at discovering and interrelating new cognitive laws such as those described above.

References

Anderson, N. H. (1981). *Foundation of information integration theory*. New York: Academic Press.

Anderson, N. H. (1982). *Methods of information integration theory*. New York: Academic Press.

Anderson, N. H. (1983). Cognitive algebra in intuitive physics. In H.-G. Geissler, H. F. J. M. Buffart, E. L. J. Leeuwenberg, & V. Sarris (Eds.), *Modern issues in perception* (pp. 229–253). Amsterdam: North-Holland.

Anderson, N. H. (1996). *A functional theory of cognition*. Hillsdale, NJ: L. Erlbaum Associates.

Anderson, N. H., & Cuneo, D. O. (1978). The height + width rule in children's judgments of quantity. *Journal of Experimental Psychology: General*, *107*, 335–378.

Brown, W., & Thomson, G. H. (1921). *The essentials of mental measurement*. London: Cambridge University Press.

Cliff, N. (1992). Abstract measurement theory and the revolution that never happened. *Psychological Science*, *3*, 186–190.

Cocco, A., & Masin, S. C. (2010). The law of elasticity. *Psicológica*, *31*, 647–657.

Corneli, E., & Vicovaro, M. (2007). Intuitive cognitive algebra of sliding friction. *Teorie & Modelli*, *12*, 133–142.

De Marchi, S. (1925). La valutazione di collettività. In [No Author] *Atti del IV Congresso Nazionale di Psicologia* (pp. 131–134). Firenze: Bandettini.

Ellis, B. (1966). *Basic concepts of measurement*. Cambridge: Cambridge University Press.

Estes, W. K. (1975). Some targets for mathematical psychology. *Journal of Mathematical Psychology*, *12*, 263–282.

Fagot, R. F., & Stewart, M. (1969). Tests of product and additive scaling axioms. *Perception & Psychophysics*, *5*, 117–123.

Falmagne, J. C. (1974). Foundations of Fechnerian psychophysics. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical* psychology (Vol. 2, pp. 127–159). San Francisco, CA: Freeman.

Falmagne, J. C. (1985). *Elements of psychophysical theory*. New York: Oxford University Press.

Fechner, G. T. (1860). *Elemente der Psychophysik*. Leipzig: Breitkopf und Härtel.

Fullerton, G. S., & Cattell, J. M. (1892). *On the perception of small differences with special reference to the extent, force and time of movement*. Philadelphia: University of Pennsylvania Press.

Gescheider, G. A. (1997). *Psychophysics: The fundamentals*. Mahwah, NJ: Erlbaum.

Gigerenzer, G., & Richter, H. R. (1990). Context effects and their interaction with development: Area judgments. *Cognitive Development*, *5*, 235–264.

Hearnshaw, J. B. (1996). *The measurement of starlight: Two centuries of astronomical photometry*. Cambridge: Cambridge University Press.

Herschel, J. F. W. (1829). Observations with a 20-feet reflecting telescope.—Third series; containing a catalogue of 384 new double and multiple stars; completing a first thousand of those objects detected in a course of sweeps with that instrument: together with observations of some previously known. *Memoirs of the Astronomical Society of London*, *3*, 177–213.

Holland, P. [Translator] (1601). *The historie of the world.* London: Adam Islip.
Luce, R. D. (1972). What sort of measurement is psychophysical measurement? *American Psychologist, 27*, 96–106.
Luce, R. D. (2002). A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychological Review, 109*, 520–532.
Luce, R. D., & Galanter, E. (1963). Psychophysical scaling. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (Vol. 1, pp. 245–307). New York: Wiley.
Marks, L. E., & Algom, D. (1998). Psychophysical scaling. In M. H. Birnbaum (Ed.), *Measurement, Judgment, and Decision Making* (pp. 81–178). New York: Academic Press.
Masin, S. C. (2008). Is self-estimated linear length linear? In B. A. Schneider, B. M. Ben-David, S. Parker, & W. Wong (Eds.), *Proceedings of the twenty-fourth annual meeting of the International Society for Psychophysics* (pp. 265–270). Toronto, Canada: International Society for Psychophysics.
Masin, S. C., & Toffalini, E. (2009). On the bisection operation. *Attention, Perception, & Psychophysics*, *71*, 1566–1663.
Merkel, J. (1888). Die Abhängigkeit zwischen Reiz und Empfindung. *Philosophische Studien, 4*, 541–594.
Narens, L. (1996). A theory of magnitude estimation. *Journal of Mathematical Psychology, 40*, 109–129.
Pfanzagl, J. (1959). A general theory of measurement applications to utility. *Naval Research Logistics Quarterly, 6*, 283–294.
Plateau, J. (1872). Sur la mesure des sensations physiques, et sur la loi qui lie l'intensité de ces sensations à l'intensité de la cause excitante. *Bulletin de l'Académie royale des Sciences, des Lettres et des Beaux-Arts de Belgique, 2e série*, *33*, 376–388.
Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. London: Routledge and Kegan Paul.
Richardson, L. F. (1929a). Imagery, conation, and cerebral conductance. *Journal of General Psychology, 2*, 324–352.
Richardson, L. F. (1929b). Quantitative mental estimates of light and colour. *British Journal of Psychology*, *20*, 27–37.
Richardson, L. F., & Ross, J. S. (1930). Loudness and telephone current. *Journal of General Psychology*, *3*, 288–306.
Scheerer, E. [Translator] (1987). My own viewpoint on mental measurement (1887): Gustav Theodor Fechner. *Psychological Research*, *49*, 213–219.
Schönemann, P. H. (1994). Measurement: The reasonable ineffectiveness of mathematics in the social sciences. In I. Borg & P. Ph. Mohler (Eds.), *Trends and perspectives in empirical social research* (pp. 149–160). Berlin: de Gruyter.
Shepard, R. N. (1966). Metric structures in ordinal data. *Journal of Mathematical Psychology*, *3*, 287–315.
Steinheil [C. A.] (1837). Elemente der Helligkeits-Messungen am Sternenhimmel. *Abhandlungen der mathematisch-physikalischen Classe der Königlich Bayerischen Akademie der Wissenschaften*, *2*, 1–140.
Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: Wiley.
Toomer, G. J. [Translator] (1984). *Ptolemy's Almagest*. London: Duckworth.
Weaver, H. F. (1946). The development of astronomical photometry. *Popular Astronomy, 54*, 211–230.
Weiss, D. J. (1972). Averaging: An empirical validity criterion for magnitude estimation. *Perception & Psychophysics*, *12*, 385–388.