TO HONOR STEVENS AND IMPROVE HIS SCALING METHODS

Gunnar Borg, PhD
Prof. em of perception and psychophysics
Stockholm University, 106 91 Stockholm, Sweden
E-mail: gbg@psychology.su.se

Abstract

This paper is devoted to honor Stevens for his great contributions for psychology. Studies are presented that I have performed over the past fifty years, starting with experiments, where I used Stevens scaling methods in some new fields. These studies would not have been possible without his new methods I will then present some criticism and the results of my work to achieve "level-anchored ratio scaling".

The immense and epoch-making importance of the scientific contributions of Stanley S. Stevens is shown in many ways. One of these is all the work inspired by his work.

Many problems in science have not been studied because of the lack of good methods. Peter Noll, the Swiss scientist who wrote a book ("Dictate über Sterben und Tot", 1984) in the form of a diary of his last year before he died from cancer and refused any therapy, makes some remarks about this. I guess that when someone finds that the time is limited, he or she may think of how many important experiences we may have had without knowing or understanding much about them, because we don't have a valid method to study them. Noll mention such things as love, happiness, sorrow and contentment. He means that modern science has an enormous ability to leave aside questions and that these then are dealt with by pseudoscience, such as astrology. – Of special importance in psychophysics is Stevens' scaling methods. These have helped studying some of the problems Noll mentions. To learn about these methods has been of major importance.

Applications of Stevens ratio scaling

Stevens showed the importance of "ratio scaling" to describe how a stimulus (S) is perceived (R), and that the simple category scales do not function well. As an example how the S-R function can deviate from a linear relation, the method of halving or doubling the perception of an S can be used. In one of my very first experiments at the end of 1950s I started from the common experience that people perceive a decrease from 100 km*h^{-1} to 50 as very much bigger than a halving. By having people do several halvings and doublings, I found that the perceived speed roughly varies with the square of the physical speed. I often use this result as a simple pedagogical example to demonstrate a psychophysical function. When I visited Bert Scharf in Boston about 25 years ago, his car was in a garage and he asked his neighbor, a physicist, to come and pick me up at the airport. He asked me about my scientific field, and what psychophysicists dealt with. I then told him about "measuring" perceptual magnitudes and demonstrated it by using the speed experiment when we were driving to the university. When we decelerated from 50 miles an hour, and I asked him to tell me when the speed had been reduced by half, he told me that when it was about 35. He was astonished about the result and became interested.

I then started to make experiments about perceived exertion in heavy ergometer work. In this field as well, perception followed a positively accelerating function with an exponent of about 1.6. I had to add an extra constant referring to a basic perceptual "noise" at rest, in a similar way as Ekman (1959) did for perception of saltness. Stevens didn't like that but I insisted when I met him during 1968. We had a very nice and rewarding meeting that lasted the whole afternoon. – I then also included one more constant, referring to the fact that not all S-R-functions start at zero stimulus intensity. For perceived exertion in walking we need to add a constant on the S-axis, since there is no increase in perceived exertion when walking very slowly. The general S-R-function should therefore be expressed in the following way (Borg, 1961, 1962, 1973): $R=a+c(S-b)^n$. The inclusion of these extra constants can easily be defended by complementary studies to determine thresholds etc., in analogy with physiological experiments, e.g., the increase of blood lactate during exercise (Borg 1962). This equation was then also proposed by Mountcastle et al (1963). My dissertation (1962) dealt with these problems and I was very pleased to be congratulated by Stevens for having opened up a new field for scientific investigations.

In the 1960s Stevens' ratio scaling methods met with some criticism. The results obtained could not be interpreted as belonging to an absolutely true ratio scale. This didn't bother me too much since the methods seemed to work well – giving results as a "semi-ratio-scale", as I later liked to name it – at least for rough determinations of differences between sensory modalities. I was then very fortunate to collaborate with an internationally leading physiologist, Yngve Zotterman. He had been working together with an ear doctor (Diamant) and they had succeeded to pick up neurophysiological responses from the taste nerve. By a freak of nature this goes through the middle ear and during surgery to improve the hearing it was possible to get responses when stimulating the taste buds with, e.g., citric acid and sugar. The test stimuli consistent of diluted solutions from very weak to very strong. I tested the patients before the surgery and the same stimuli were then used in the physiological test. A very high correlation was found between the perceptual and the physiological responses. S-R-functions with very similar intramodal exponents were found, for citric acid about 0.67 and for sugar about 1.0. This was, probably, the first time such a direct neuropsychophysiological study on human beings could be performed and it gave strong support for the value of "ratio scaling" (Borg et al. 1967).

Fifty years ago I was doing some research and teaching, but also applied work as psychologist at the university hospital. I have since then for more than 30 years tried to combine scientific work with applications in medicine, ergonomics and human factors, and sports. I think this combination has been fruitful both for the research and for the applications.

At the hospital category scales were used. They gave rough information of levels of intensity. That a patient told me that a feeling was twice as strong as another one was interesting, but it was necessary to know how strong it was. When testing lumber workers I found that when they used the word "heavy" the perception was about three times stronger than when they said it was "light". So I became interested in a new quantitative semantics, using Stevens' ratio scaling as method and adjectives and adverbs as stimuli. My conclusion was that by the results of such studies "linguistic expressions can be fitted into a subjective ratio scale so that exact mathematical relations between them be obtained" (Borg, 1962). Studies were then performed to get information about the meaning of the concepts "interpretation" and "preciseness", meaning – for a certain quality – the level of intensity in the natural perceptual range, and the degree of interpersonal agreement. In one study 37 expressions were used and their meaning in the form of positions (mean values) and the preciseness (coefficients of variation) determined (Borg and Lindblad, 1976).

Differences in ratings depend upon several factors such as perceptual inequalities, rating behaviour and context. It's a difficult task to give ratings according to ratio scaling. This was shown by Borg, G. and Borg, E. (1990). Ratings of sizes of areas were used as stimuli and ME with a standard and modulus as psychophysical method. Three groups participated: students in mathematics, in pedagogy and in home economics. The exponents differed between the groups, from 0.86 (math), 0.83 (ped) to 0.62 (home econ.). The students of home econ. used fewer low numbers and not so many high numbers. The students of mathematics used the greatest range of numbers. The study shows that a big variation in responses may be obtained depending upon how competent the observers are.

It's impossible for people to give responses with numbers in the form of true mathematical concepts. We use our number conceptions, and to state that "ratio scaling" can give responses on a ratio scale can never be correct. This does, however, not mean that "ratio scaling" is useless for the purpose of determining S-R-functions. A concept that I have liked to use is "semi-ratio". A "semi-ratio scale" (Borg, 2002) then means that results obtained can be treated as if they belong to a true ratio scale. The best correlations with physiological measures and predictions to performances can then be obtained, than otherwise would be possible.

It's commonly hold that a category scale only possesses ordinal metric properties. However, in spite of this, a typical relation to a ratio scale has been obtained (Stevens and Galanter, 1957). Eisler (1962) found that by taking the logarithm of the ratio scale a linear relation was obtained. This means that some rating scales may partly be characterized as "semi-log scales". Since a log-function is the inverse of an exponential function it should then also be possible to "stretch out" the category scale by an "exponential transformation" to a "semi-ratio scale". This was proposed by Borg, G and Borg, P. (1987) and tested with good results.

Studies with Stevens' ratio scaling have been performed about the perception of difficulty of mental tasks. The psychometric methods are using solution frequencies and express the degree of difficulty in z-values. When I was testing patients I found that sometimes they pondered too long to solve a certain item (in a time-limited test), but that they then could solve more difficult problems. They didn't get time to solve all the problems that they could have done. When I analyzed the different items my impression was that some items seemed to be more difficult, than they were. A study by Borg and Forsling (1964) gave an interesting result. Nine items from a test of reasoning ability were chosen with z-values from -1.2 to +1.2. Magnitude estimation was used and the item with z=0.0 was used as standard with modulus "10". The median ratings varied from 5.1 to 19.6 and corresponded well with the z-values. There was, however, a great variation of the dispersions. As an example item No. 6 had a coefficient of variation of 74; No. 7:35; No. 8:89; and No. 9:46. My immediate impression of some items being ambiguous was thus confirmed.

A series of studies were then performed on many different kinds of mental tests. High correlations between ratings and z-values were commonly found (0.80 – 0.90). The psychophysical method added an important measure; viz. dispersions for each item! The perceived agreement of the difficulty across subjects varied sometimes more than expected (Borg et. al., 1971). In some experiments an absolute zero on the z-scale was also estimated. By extrapolating the obtained S-R functions (with the z as S), down to the zero R-value, the intersection on the x-axis, a psychometric zero could be estimated. These were of the magnitude -4z to -5z. The experiments showed an interesting psychophysical – psychometric possibility (Borg, 1977).

Drawbacks with Stevens' ratio scaling

There are many studies that show the importance of Stevens' ratio scaling methods. They function in a similar way as those in natural sciences, where an arbitrary and "meaningless" intensity is chosen as a unit. A main difference to scaling in physics is that the unit is universal, while the psychophysical unit is subjective. It's private and may differ rather much between people. During previous ISP-meetings we have discussed this problem. During the meeting in Stockholm I cited Quine (1987) and his statement that in science we cannot talk about "bigness" only about "biggerness" because there are no boundaries. My comment to this was that we have our boundaries, e.g., from the exertion at rest to that during a maximally strenuous work. We need to identify our boundaries and use these for different modalities to be able to identify levels of intensity (Borg, 1992). As Lawrence Ward so rightly has pointed out, Quine is wrong not only for all living creatures, but also for other things such as technical apparatuses.

A physical measure is "meaningless" if we don't know what class of objects or events we are dealing with. We cannot say that 1 m is long or 1 kg is heavy if we don't know the class and context. Numbers don't refer to any special intensity. The advantage with Stevens' ratio scaling to obtain valid S-R-functions then also involved and caused a disadvantage with regard to the possibility to obtain an "absolute" intensity. Stevens enforced the idea that it's necessary for "magnitude estimation" to function as a ratio scaling method that the subjects should be permitted to use numbers as they preferred, "without any restrictions". He then also condemned the category scales with their small and restricted range of numbers (Stevens, 1971).

When using "ratio scaling" it's not possible to get a direct rating (direct response, immediately possible to interpret) denoting a level of intensity. To enable interindividual comparisons I proposed a working model, the "range model", according to which perceptual magnitudes are approximately equal for different people at a highest possible stimulation. The subjective, dynamic range should then be very similar (Borg, 1961, 1990, Sagal and Borg, 1993). The implications of the model has been tested in several studies and found to work well (see Borg, 1998). The model was also used in the study on taste perception. The differences in exponents between perceived sourness and perceived sweetness was then "explained" by the big difference in stimulus range between these two qualities, since the perceptual range was chosen to be equal. As we then expressed it: "The reason for this may be the fact that the range of discrimination is so much greater for citric acid than for salt and sucrose" (Borg et. al. 1967). This model concern biologically natural ranges and not experimentally manipulated ones as Parduccis (1965) range-frequency model does. For intermodal comparisons the range model was used in a very important study by Teghtsoonian (1971), which showed that most of the differences in exponents could be explained by the differences in S-ranges. To me this is the best validation and "explanation" of the variation of exponents in Stevens' power law.

Category-Ratio (CR)-scaling

Stevens' ratio scaling has the great advantage of enabling determinations of relative S-R-functions. However, the simple category scales have the advantage of giving direct levels of intensity. To combine the advantages of these two types of scaling methods, the category (C) – ratio (R) scales were constructed. The aim was to get a "level-anchored ratio scale". Several demands had then to be fulfilled. I will now just summarize some of these: The Range Model. The size of the natural, subjective dynamic range. The choice of verbal anchors. One main anchor as "a fixed star" (a schematized conception). The positioning of the anchors to obtain congruence between numbers and anchors. The avoidance of end effects. The possibility to get direct responses. Common psychometric demands referring to internal consistency, validity, applicability etc. See "A new generation of scaling methods: Level anchored ratio scaling", Borg, G and Borg, E. (2001).

Several empirical tests have been performed in the scale construction and validations. Similar S-R-functions should be obtained as those obtained with Stevens' methods. It was then important that the scale should make possible interindividual comparisons with regard to levels of intensity. Several versions of CR-scales have been constructed. The most common are the CR10-scale and the CR100, also called the centiMax-scale. That magnitude estimation, "ME", doesn't function well for differential purposes has been shown in studies by Borg. That this also is true for absolute magnitude estimation, "AME", has been shown by, e.g., Berglund (1991), Borg, E and Borg, G (2002).

Many experiments have shown the effect of "context", e.g., the conditions and design of an experiment, what range and frequency of stimuli are used. To me the importance of the context is rather much overestimated. It's very easy to arrange experiments to show the effect of the context by rearranging ranges and expose people to many repetitions that they will adapt to. With common clinical tests, the situation and the design are, however, very similar from lab to lab. When I have lectured in hospitals in European countries, in the USA, Brazil or Japan, I have been surprised how similar things are. The "Borg-scales" (not my designation) are used all over the world and millions of people are exposed to one of my scales each year. They are of special value in clinical diagnostics. Ratings of dyspnea with my CR10 scale, as just one example, have become common. In studies by O'Donnel et. al (1997) very high correlations have been found with physiological measures of pulmonary functions (correlations about 0.70-0.80). There are many such examples that show the value of CR-scaling and its advantage over just ratio scaling. For most kinds of perceptions and feelings (including symptoms of medical importance), it's thus possible to determine individual differences of both theoretical and applied interest.

References

Berglund, B. (1991). Quality assurance in environmental psychophysics. In S.J. Bolanowski & G.A. Gescheider (Eds.), *ratio Scaling of Psychological Magnitudes – In Honor of the Memory of S.S. Stevens.* Hillsdale: Erlbaum; pp. 140-162.

Borg, E. and Borg, G. (2002). A comparison of AME and CR100 for scaling perceived exertion, *Acta Psychologica*, Vol. 109, issue 2, pp. 157-175.

Borg, E., Kaijser, L. (2006). A comparison between three rating scales for perceived exertion and two different work tests. *Scand J Med Sci Sports.16,* pp. 57-69.

Borg, G. (1961) Interindividual scaling and perception of muscular force*. Kungliga fysiografiska sällskapet i Lund förhandlingar, 12, 31,* pp. 117-125*.*

Borg, G. (1962). Physical performance and perceived exertion. *Studia Psychologica et Paedagogica,* Series altera, Investigationes XI. Lund, Sweden: Gleerup.

Borg, G., and Forsling, S. (1964). A psychophysical study on perceived difficulty. *Educational and Psychological Research Bulletin*, the University of Umeå, No. 1, pp.6

Borg, G., Diamant, H., Ström, L., and Zotterman, Y. (1967). The relation between neural and perceptual intensity: A comparative study on the neural and psychophysical response to taste stimuli. *Journal of Physiology*, *192*, pp. 13-20.

Borg, G., Bratfisch, O., and Dornic, S. (1971). On the problems of perceived difficulty. *Scandinavian Journal of Psychology*, 12, pp. 259-260.

Borg, G. (1973). Perceived exertion during walking: A psychophysical function with two additional constants. Report from the Institute of Applied Psychology, the University of Stockholm, No. 39, pp. 10.

Borg, G., & Lindblad, I. (1976). The determination of subjective intensities in verbal descriptions of symptoms. *Reports from the Institute of Applied Psychology*, Stockholm University, No. 75.

Borg, G. (1977) Aspects subjectifs de la charge physique et mentale. *Le travail humain, tome 40, no 2,* pp. 225 à 232.

Borg, G. & Borg, P. (1987). On the relations between category scales and ratio scales and a method for scale transformation. *Reports from the Department of Psychology*, Stockholm University, No. 672.

Borg, G. (1990). A General Model for Interindividual Comparison. In: Jm. W. Baker, M.E. Hyland, R. van Hezewijk, & S. Terwee (Eds.). *Recent Trends in Theoretical Psychology, Volume II.* (pp. 439-444). New York: Springer-Verlag.

Borg, G., & Borg, E. (1990). Psychophysical judgment of size and their use to test rating behavior. *Reports from the Department of Psychology,* Stockholm university, No. 717.

Borg, G. (1992). A ”fixed star” for interprocess comparisons. In: G. Borg, & G. Neely (Eds.), *Fechner Day 92. Proceedings of the Eight Annual Meeting of the International Society for Psychophysics* (pp. 41-45). Stockholm: Department of Psychology, Stockholm University

Borg, G. (1998). Borg’s Perceived Exertion and Pain Scales. Champaign, IL:Human Kinetics.

Borg, G. and Borg, E. (2001). A new generation of scaling methods: Level-anchored ratio scaling. *Psychologica.* 28**,** pp. 15-45.

Borg, G. (2002). Basic principles in ”Level-Anchored-Ratio-Scaling”. In: J. Da Silva, E. Matsushima, N. Ribeiro-Filho (Eds.) *Fechner Day 2002. Eighteenth Annual Meeting of the International Society for Psychophysics* Vol. 18. (pp. 266-271). Rio de Janeiro: Editora Legis Summa Ltda.

Eisler, H. (1962). Empirical test of a model relating magnitude and category scales. *Scandinavian Journal of Psychology*. *3*,pp. 88-96

Ekman, G. (1959). Weber's law and related functions. *Journal of Psychology*, *47,* 343-352.

Mountcastle, V. B., Poggio, G. F., & Werner, G. (1963). The relation of thalamic cell response to peripheral stimuli varied over an intensive continuum. *Journal of Neurophysiology, 26,* pp. 807-834.

Noll, P. (1984). Dictate über Sterben und Tot. Pendo Verlag AG, Zürich.

O’Donnel, D., Bain, D. and Webb, K. (1997). Factors Contributing to Relief of Exertional Breathlessness during Hyperoxia in Chronic Airflow Limitation. *Am J Respir Crit Care Med*. *Vol 155,* pp. 530-535.

Parducchi, A. (1965). Category judgment: A range-frequency model. *Psychological Review, 72,* 407-418.

Quine, W. V. (1987). Quiddities. Harvard University Press.

Sagal, P., & Borg, G. (1993). The range principle and the problem of other minds. *Brit J Phil Sci, 44,* pp. 477-491.

Stevens, S. S. (1971). Issues in psychophysiological measurement. *Psychological Review, 78,* pp. 426-450.

Stevens, S. S., & Galanter, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, *54,* 377-411.

Teghtsoonian, R. (1971). On the exponents in Stevens´ law and the constant in Ekman´s law. *Psychological Review, 78,* pp. 71-80.