

COMMENTARY: DIRECT VERSUS INDIRECT SCALING: THE  
GNASHING OF PSYCHOPHYSICAL WORLDVIEWS

*A.V. CARDELLO*

The advantages and disadvantages of choice-based or indirect methods of scaling versus direct methods have been argued passionately for over a century. Both positions have proponents and detractors; both have theoretical arguments and empirical data to support them; and both have the potential to provide insights into important product development issues. What is also clear is that they constitute two distinct approaches to data collection and analysis that, under certain conditions, produce discordant data, but that from a practical perspective, rarely result in information that is logically contradictory or that would result in an entirely different course of action regarding the product development issue at hand. Thus, it would seem that any fruitful discussion should not focus on which methods are correct or incorrect, but on the practical and theoretical differences between the two and the reasons why scientists of different persuasions choose one or the other.

In his introduction, Moskowitz traces the “indirect” approaches to Fechner. In keeping with the zeitgeist, Fechner believed that sensations were internal events that could not be measured directly. However, it must also be noted that in the years leading up to his groundbreaking work in psychophysics, Fechner experienced a spiritual and religious awakening that turned him away from materialism and to a philosophy that sought to uncover the spiritual meaning in the world. His philosophical conversion led him to the belief that mind and matter were two aspects of the same reality and that everything in nature was part of a single consciousness. It was this panpsychic philosophy that drove Fechner’s intellectual thinking and that led him to apply his training in physics to the development of a new science that he hoped would demonstrate the oneness of the human psyche and the physical world. I make this point as a prelude to later comments about how broader psychological and philosophical worldviews influence the methods and approaches that psychophysicists find logical, rational, and/or intellectually appealing.

In the same way that Fechner believed that sensation magnitude could not be measured directly, so too did Thurstone assert that “there is no possibility of recording experimentally in any direct way the discriminial processes that correspond to a series of stimuli” (Thurstone 1927, p. 370). Thurstone’s use of the term “discriminal process” instead of “sensation” underscored his fundamental belief that sensations were not directly measurable and that even the word “sensation” should not be used to refer to what is being quantified. In his opening assault on Thurstonian scaling, Moskowitz states “the reality of the situation is that choice behavior does not provide a metric for sensory magnitude, which is a necessity in developmental work.” This is a bit of hyperbole, since choice-based methods do provide “a metric,” albeit neither

a direct nor absolute one. While Thurstone admitted the reality of sensation magnitude, he made no assumptions about it – “I shall not assume that sensations, or whatever the identifying and discriminating functions may be called are magnitudes. It is not even necessary for the present argument to assume that sensations have intensity” (Thurstone 1927, p. 368). The point here is that the “lack of metrics” that Moskowitz bemoans is not so much a *failing* of Thurstonian approaches, as it is a direct *consequence* of Thurstonian assumptions (or lack thereof). To me, what is most peculiar about Thurstone’s views is that, while he felt it was unnecessary to make the rather safe assumption that sensations had magnitude, he was quite willing to make the far more risky assumption that, given equal dispersion, equally often noticed differences are psychologically equal.

In making the “lack of metrics” argument, Moskowitz raises the example that if two stimuli are completely discriminable (100% choice of one stimulus over the other) then the Thurstonian system “totally falls apart.” This too is overstated, since one need only insert intermediate stimuli that are not completely discriminable from either of the extreme stimuli to enable the construction of an intensity scale that includes all stimuli. Of course, the necessity to add stimuli that are not of practical interest to the developer highlights the very real problem that Moskowitz raises, i.e., that choice-based methods are difficult to execute due to the large number of paired comparisons that are required. This criticism is indisputable and stands as the greatest impediment to a still wider adoption of choice-based methods in business and marketing.

Another practical problem of Thurstonian approaches that Howard offers is that, for food and other consumer product applications, choice-based methods never compare the products simultaneously. Instead, one sample is compared after consumption (or use) to another sample that was consumed previously and that is now simply a memory to the consumer. This is a valid point. Even Thurstone himself felt that the use of paired comparisons for food testing was unworkable. Thurstone worked under contract for the Quartermaster Food and Container Institute in Chicago after WWII. In a paper published in 1950 in the Proceedings of the Second American Meat Institute Conference on Research, Thurstone stated:

In some types of psychophysical experiments one can present all of the stimuli in pairs, so that each subject makes a comparison of each pair of stimuli . . . In practice, such a procedure is almost out of the question in dealing with taste.

He went on to say:

The best procedure is probably to have a fairly large number of subjects and to ask each subject to sample each specimen only once . . . he (the

subject) may be asked to allocate the specimens to a set of, say, 10 steps. These might be numbered from 1 to 10, and he might be asked to let number 1 represent the most disagreeable, while number 10 represents the most agreeable, taste. These 10 steps would really represent intervals on a subjective scale of taste preference . . . (Alternatively, the subject) states his degree of preference in terms of one of a number of short descriptive phrases which are assigned to the successive intervals. There is no assumption that these successive intervals of the scale are in any sense equal . . . For convenience, these descriptive phrases may be denoted by numbers or letters. (Thurstone 1950, pp. 86–87).

These statements make it clear that Thurstone felt that the use of paired comparisons in food research was impractical. In addition, his comments provide an interesting historical insight into a critical period in the evolution of consumer hedonic testing. It should be remembered that Thurstone wrote these comments during the time that he worked at the Quartermaster Institute with Peryam and Pilgrim. Thus, it is of some significance that we see reflected in his comments the basic elements of the nine-point hedonic scale that was under development there, and that was to be published shortly. It is ironic that the nine-point hedonic scale, now the most commonly accepted direct measure of liking/disliking, would soon replace the laborious paired comparison methods that Thurstone championed and that had dominated earlier research in our field.

One practical point that Howard does not raise in his essay, but that is a real strength of choice-based methods, especially to multinational companies, is their reduced reliance on spoken and written language. Direct measures of scaling require the use of rating scale labels that must be generally understood by subjects in order to make valid comparisons of either sensory or hedonic judgments. Although the use of unstructured scales or magnitude estimation can minimize language differences, choice-based methods eliminate the problem by substituting a behavioral choice of one sample (or concept) over another. The development of best-worst (maximum difference) scaling by Louviere and collaborators is a practical example of the extension of Thurstonian methods to meet the specific need to compare preferences in multinational markets.

While Moskowitz's essay serves as a general attack on indirect methods, he focuses a great deal of his discussion on choice-based conjoint measurement. Here he sees a particularly disturbing problem because of the recent resurgence in choice-based approaches. This trend is certainly surprising, given the practical problem of incorporating the number of pairs required for choice-based conjoint, but also because choice-based approaches to conjoint are relatively inefficient for identifying consumer segments, due to the difficulty of estimating utilities for individuals.

One of the major reasons cited by Howard in his attempt to explain the popularity of choice-based methods in business is their “intuitive appeal,” i.e., the act of choosing in a paired comparison test or choice-based conjoint mimics the ultimate behavior of interest – the choice of products in the marketplace. Of course, the similarity between the two situations is entirely superficial due to the large contextual differences in the situations. Moreover, it can be argued that choice behavior fails to mimic actual trial and use, which is usually done at home and without a comparison product. Although, initial choice occurs in the marketplace (without tasting), decisions about continued use are made at home – monadically. Nonstatistical contextual differences between choice-based and direct methods will, in part, account for any observed differences in results, independent of differences in the statistical models.

While Moskowitz states that the “intuitive appeal” of choice-based methods wanes when it comes to the complex task of converting the paired comparison data to scale values, he is surprisingly understated when it comes to pointing out the obvious and logical intuitive appeal of direct scaling, i.e., that if you desire to know how strongly a person perceives a sweet taste to be or how much he/she likes a product, you can simply ask them directly. It seems entirely counterintuitive to set up a set of choices for the subject and to then try to determine their sensory or hedonic experience(s) by working backwards from their choices to construct a “scale.” This, to me, is the fundamental puzzle surrounding the propensity to use choice-based methods over direct scaling.

Moskowitz’s essay moves from the practical issues of scaling to the more important issue of the contribution that each approach has made to the advancement of science. His argument that “no substantive scientific laws or at least generalities appear to have emerged from paired comparison studies” is controversial at best. The overarching concern that he expresses reflects his desire for schema over empirics, for theory-driven research over data-driven application and for the nomothetic pursuit over ideographic testing. It is a theme that Howard has raised repeatedly throughout the years, and one that deserves particular attention in a field like ours, where the demand for practical testing is so great that the pursuit of causal mechanisms and general principles may be lost in the daily rush to solve immediate product development problems. Howard gives credit here to the research by Ennis and others that attempts to fully develop the promise of Thurstonian approaches, but he questions whether these efforts will “bear fruit.” Whether the fruit of this research will be (or already has been) born, it is clear that these efforts *do* serve the advancement of science. Science proceeds best when existing theories are questioned, alternative theories proposed, critical experiments undertaken and improved theories generated. While Howard chides those in

business who use choice-based methods out of “inertia, reflexivity, fuzzy thinking, convenience and laziness,” in fairness, we must applaud those who use these methods in an intelligent and systematic attempt to construct a more elaborate theoretical and empirical foundation that may be compared to other approaches to scaling. Only in this manner can we avoid the potential for intellectual laziness to produce unquestioned acceptance of our *own* personal point of view.

At the end of his essay, Howard enumerates several reasons why “the archaic paired comparison methods” still hold traction in business applications. Among these, he cites inertia of the past, lack of results-accountability, reflexive demands for head-to-head measures of superiority, fuzzy thinking, expediency and a decline in statistical knowledge. The one reason among these that seems obvious is the demand for head-to-head measures. I think that some of the other reasons he cites reveal a sense of frustration on his part – a frustration born of the fact that the use of choice-based methods flies in the face of his fundamental worldview, one that is influenced by a personal psychology and philosophy of measurement that was molded by his training and experiences as a student of S.S. Stevens. Now, I don’t profess to have knowledge of Howard’s broader psychological or philosophical views, but it is clear from a career of his writings that he believes deeply in the validity, and indeed, *rightness*, of direct approaches to the measurement of sensation. (I use the word “rightness” deliberately for its teleological and even theological overtones.) In the same way, I have no doubt that the supporters of Thurstonian methods possess their own righteous belief in the validity of their position.

But, must one of these two psychophysical perspectives be “right” and the other “wrong”? Can one look to other situations in science from which to draw conclusions about which of two different explanations of reality bears truth?

In its essence, the problem is similar to the one that faced physicists in the early 20th century in dealing with the dualistic nature of light and atomic structures (waves versus particles). Here too there existed theoretical and empirical data to support two quite contrary views, and so too did methods of measurement produce differences in the empirical data under specified conditions. The potential schism of theory, data and methods in physics was resolved by the conceptualization of a critical principle of science – complementarity. As articulated by Niels Bohr, “evidence obtained under different conditions and rejecting comprehension in a single picture must, notwithstanding any apparent contrast, be regarded as exhausting all well-defined information about the atomic object” (Bohr 1963, p. 6). This principle has been restated as “evidence obtained under different experimental conditions should be equally valid in explaining the phenomenon in question” (Dzendolet 1969, p.

426), and has been proposed as a solution to other, more general, controversies in psychophysics (Baird 1997). The application of this principle to the controversy over choice-based versus direct methods of scaling results in a parsimonious solution to what often has been an acrimonious debate over the validity and usefulness of these methods. The complementarity solution requires us to admit the validity of *both* choice-based and direct methods of scaling. It requires us to accept that the data obtained by the two methods provide *complementary information* regarding the hypothetical construct of interest, and lastly, it offers the researcher a *choice* between the two methods, based on the advantages/disadvantages of each that apply to the application at hand.

If one accepts this principle, the fascinating question still remains as to why some scientists adopt one approach to measuring sensory experience almost exclusively, while other scientists adopt the opposing approach. Here our discussion must necessarily return to broader philosophical viewpoints that mold the personal paradigms of psychophysicists and that influence our choice of methods. To paraphrase Bertrand Russell – science guides us with what we know; philosophy guides us with what we don't know.

Just as Fechner evolved his methods and study of psychophysics out of a strongly held personal philosophy, Howard, through his writings, also reveals a strongly held worldview, i.e., that sensations can be measured directly. Thurstone, Stevens and the contributors to this discussion, indeed, *all* working psychophysicists have their own personal philosophy and scientific paradigm to guide their thinking and their approach to psychophysics, especially in cases where the “truth” has yet to take sides.

As previously suggested, Howard's worldview was likely shaped by the intellectual discourse occurring in the corridors of William James and Memorial Halls. Others come to direct scaling through other paths. For me, it is rooted in the epistemological primacy of sense data, combined with a behaviorist tradition, wherein overt behavior is the primary datum, and language is merely behavior under the control of a shared history of stimulus-response contingencies. In this latter view, the words and numbers that people emit are a valid and critical source of data for inferential analysis. These views, combined with an overarching humanistic philosophy that fosters confidence in the abilities of human beings, trust in the veracity of their self-reports, and skepticism regarding truths to be found beyond the observable, set the stage for the appealing Stevensonian notion – that humans can directly report the magnitude of their phenomenological experiences using words, numbers or other behavioral responses, and that no further manipulation of these reports is needed in an attempt to uncover some “truer” underlying scale of sensation.

In the words of one who eloquently articulated the relationship between the practices of science and the viewpoints of philosophy, “Paradigms provide

all phenomena, except anomalies, with a theory-determined place in the scientist's field of vision" (Kuhn 1996, p. 97).

Armand V. Cardello  
*Sensory and Consumer Sciences*  
*Science and Advanced Technology Directorate*  
*U.S. Army Natick Soldier Center*  
*Natick, MA 01760-5020*



### REFERENCES

- BAIRD, J.C. 1997. *Sensation and Judgment: Complementarity Theory of Psychophysics*. Lawrence Erlbaum Associates, Mahwah, NJ.
- BOHR, N. 1963. *Essays (1958-1962) on Atomic Physics and Human Knowledge*. Interscience Publishers, New York.
- DZENDOLET, E. 1969. Basis for taste quality in man. In *Olfaction and Taste* (C. Pfaffmann, ed.) pp. 420-427, Rockefeller University Press, New York, NY.
- KUHN, T.S. 1996. *The Structure of Scientific Revolutions*, 3rd Ed., Chicago University Press, Chicago, IL.
- THURSTONE, L.L. 1927. Psychophysical analysis. *Am. J. Psychol.* 38, 368-389.
- THURSTONE, L.L. 1950. Methods of food tasting experiments. In *Proceedings of the Second Conference on Research, American Meat Institute - Council on Research*, pp. 85-91, University of Chicago Press, Chicago, IL.