

---

# Comment on Trendler's (2019) "Conjoint measurement undone"

Theory & Psychology  
2019, Vol. 29(1) 129–137

© The Author(s) 2018

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0959354318815767

journals.sagepub.com/home/tap



**David H. Krantz**

Columbia University

**Thomas S. Wallsten**

University of Maryland

## Abstract

Most of Trendler's (2019) article, "Conjoint measurement undone," seems wrong to us. We explain why we disagree completely with two of his assertions: (a) that cardinal measurement scales are absent in psychology and (b) that psychology has stagnated. We share three of his other concerns, but not his perspectives on them or the supposed links among them. These three points are: (a) fewer applications of additive conjoint measurement than initially expected, (b) flaws in the practice of statistics, and (c) need to improve the culture of replication in psychology. We provide our views on these points and also note two distinct strands in the foundational analysis of measurement—one derived from geometry, the other from probability. Trendler completely overlooked the latter.

## Keywords

conjoint measurement, measurement foundations, psychophysics, replicability

Most of Trendler's (2019) article, "Conjoint measurement undone," seems wrong to us. We structure this comment around three points where we do share his concerns, although not his reasoning or linkage of the three issues, and mention two broad assertions with which we disagree emphatically. Our points of shared concern are: (a) conjoint measurement has had much less impact on empirical research than initially expected, (b) statistical practice is often flawed in psychology and other research, and (c) psychology should strengthen its replication culture.

---

## Corresponding author:

Thomas S. Wallsten, Department of Psychology, University of Maryland, 4094 Campus Drive, College Park, MD 20742, USA.

Email: [tswallst@umd.edu](mailto:tswallst@umd.edu)

Our emphatic disagreements are with two statements: (a) “no interval or ratio scales have been established in psychology, neither by conjoint measurement nor by any other means” (p. 101) and (b) “stagnation really characterizes psychology as an empirical science” (p. 101).

We begin with these two major disagreements. Then we turn to the three shared concerns, explaining our point of view about each—different from Trendler’s. We conclude with a brief perspective on foundations of scientific measurement, pointing out two bases for measurement, one geometric (for both physical and psychological measurement), the other probabilistic (used mostly in psychology).

### **Interval or ratio scales abound**

The assertion that there do not exist interval or ratio scales in psychology ignores the fact that conjoint measurement has been used successfully to establish such scales (e.g., Falmagne, Iverson, & Marcovici, 1979; P. E. Green & Srinivasan, 1978; Wallsten, 1976). In addition, alternative, more practical measurement procedures, such as those of *functional* measurement (Anderson, 1970) or computer search for best fit within a class of models (e.g., Kruskal, 1965), also yield interval scales, though these methods generally involve additional assumptions (often not made explicit by inclusion in an axiom system).

A graver problem with Trendler’s (2019) first assertion is that he ignores the use of statistical models for choice frequency as a principal basis for psychophysical scales (e.g., Fechner, 1860; Luce, 1959; Thurstone, 1927; and cf. Chapters 4 and 17 of *FOM*).<sup>1</sup> He also ignores the large literature on measurement of perceived color summarized in Chapter 15 of *FOM*. This includes 19th and 20th-century contributions by many physicists, including Young, Grassmann, Maxwell, Helmholtz, Wright, Judd, Stiles, and Wyszecki. Lastly, he ignores the models of Shepard, Tversky, and others for perceived distance or dissimilarity, based on geometry or on graph theory (discussed in Chapters 12–14 of *FOM*).

We note that Trendler’s view of measurement scales ignores material in Chapters 4, 5, 7, 8, 10, and 11–22 of *FOM*, much of which is relevant to his assertions.

### **Psychology is advancing (despite flaws)**

Trendler’s (2019) claim that psychology is stagnant seems to rest narrowly on criticisms of item-response models by Michell (1999) and others. His broad assertion ignores the many new facts and concepts developed in psychology across the past 70 years.

Psychological research in the 20th century led to many conceptual changes related to major social issues, e.g., child-rearing practice, racism, and effects of the community on individuals. A list of valuable conceptual developments since 1950 would be long indeed. We mention three especially close to our hearts.

Signal detection theory (D. M. Green & Swets, 1966), jointly developed by engineers and psychologists during World War II, has had a continuing and profound effect on many aspects of theoretical and applied psychology. This broadly applicable approach has altered thinking about psychophysics, recognition memory, subjective and quantitative forecasting, judgment and decision-making, medical diagnosis, and more. It

provides important examples of measurement based on statistical models for choice frequencies (mentioned in the previous section).

Progress in understanding human learning and memory has been so sustained and rich as to defy summary. A starting point was the discovery of the constructive nature of memory and importance of schemas (Bartlett, 1932). Later research identified sensory, short term, long term, and working memory storage (Atkinson & Shiffrin, 1969; Cowan, 2008; Schurgen, 2018); distinguished implicit from explicit memory (e.g., Schacter, 1987); and episodic from semantic memory systems (e.g., Horner, 1990). We are now beginning to understand the neural basis of encoding, storage, and retrieval processes (Moscovitch, Cabeza, Winocur, & Nadel, 2016).

Research on decision-making has shown that accurate accounts of individual choice, stable social cooperation, and active engagement in learning and work each require new concepts. The idea of *constructed preference* was pioneered by Slovic (1995), with contributions from many others, including Tversky and Kahneman (1981). The concept of *altruistic punishment* was introduced by Fehr (see Fehr & Gächter, 2002). The distinction between promotion and prevention goals and importance of *fit* between goal orientation and task type came especially from Higgins (2005).

Other active psychologists would undoubtedly present rather different lists of exciting new findings and important new concepts.

Next we turn to our points of partial agreement with Trendler's (2019) article.

## Disappointment with conjoint measurement

Trendler correctly cites the review by Cliff (1992) and book by Michell (1999) in regard to the low number of examples in which conjoint measurement has been used to establish interval scales in psychology. Although we experienced this disappointment, in retrospect, our expectation should never have been so high. Note that successful measurement in geometry and physics long predated the modern development of formal axiomatic bases for them (Helmholtz, 1868; Hilbert, 1899; Hölder, 1901). What axiomatization does is reveal the implicit assumptions underlying successful measurement. In addition, proofs of representation and uniqueness theorems show why the procedures work when the axioms are approximately correct. Moreover, when a particular measurement representation is wrong, the falsity of a particular axiom is often diagnostic. For example, additivity of utility fails when the *joint ordering* of two factors that influence choices varies with variation in the level of a third factor complementary to one of those two. For another example, the positivity axiom for concatenation of velocity fails when one object is a beam of light.

Conjoint measurement provides insights regarding some types of measurement in psychology and economics by formalizing the conditions under which tradeoffs among two or more variables that affect behavior follow an additive (or other polynomial) law. Proofs of theorems in *FOM* usually suggest measurement procedures involving calibrated "measuring rods" (the standard sequences, on which Trendler focuses). However, as mentioned above in connection with the abundance of interval and ratio scales, alternative measurement procedures are often used instead.

There are two lessons to be learned here. The first is that the primary benefit of axiomatization is to convert hidden but important implicit assumptions into clear and explicit

ones. The second lesson is that the measurement procedures often differ considerably from constructions embedded in proofs of representation and uniqueness theorems. We reject Trendler's conclusion that there are no interval or ratio scales in psychology, both because it is obviously incorrect (as detailed above) and because the two lessons we cite are reinforced both by the history of axiomatic analysis (especially in geometry) and by current measurement practice in psychology.

## Flaws in statistical practice

We are horrified by much of the statistical practice in psychology and other research. But so are many other critics. While each critic "knows" how practice should change, the recipes don't agree. Proposals include: replacing null-hypothesis tests with confidence intervals (perhaps even banning null-hypothesis tests!); strongly emphasizing the centrality of statistical models in inferential reasoning, increasing sample size and thus statistical power, and replacing "frequentist" methods with "Bayesian" ones. Hardly anyone follows Trendler (2019; or Stevens, 1946) by asserting that development of interval-scale measurement is a prerequisite for statistical analysis. Various reasons justify neglecting scale type as an important factor. One point is that statistical inference can often rest on bi- or multi-nomial models for frequency counts, which avoids the question of permissible scale transformations. Experience shows that results obtained from analysis of count data scarcely ever differ appreciably from those obtained by treating ordered categories, coded as integers, as though they constitute an interval scale.

It is therefore difficult for us to see why measurement problems in psychology (if any) should be singled out as a source of difficulties with statistical inference.

## Replication and scientific progress

The matter of replication, or failure thereof, is a complicated one. Replication is abetted by statistical thinking, but not closely tied to it. It was important in science long before the burgeoning of statistics in the late 19th and the 20th century. We admire the culture of physics, which leads most new findings to be replicated promptly by other laboratories. But it is important to emphasize that replication often requires equipment that may or may not be commonplace. Roentgen's discovery of X-rays used only an induction coil, a vacuum tube, cardboard for shielding, and a photographic plate; following his report (January 1, 1896) it was replicated within a month in many European and American laboratories (Pais, 1986, pp. 37–39). Tversky and Kahneman (1971) used a brief questionnaire and an available pool of human respondents to discover that *subjective* binomial sampling distributions do not vary with stated sample size. One of us replicated this using 50 students (in a graduate statistics class) within weeks after receiving their draft manuscript and we have both since replicated it several times in classroom settings.

The culture of replication depends on feasibility, habit of mind, and typical sizes of reported effects. The *apparent* size of an effect depends on its "true" size perturbed by both "random" and "systematic" error (the latter varies among labs, but does *not* converge to zero within any single lab as  $N \rightarrow \infty$ ). Sometimes the sum of true effect plus random plus systematic error yields an apparently large finding. It excites attention, but

later seems to shrink. In the opposite case, where the *apparent* effect size is small or zero, but (later) an important effect is found, the earlier experiments may be criticized unjustly as lacking statistical power. In fact, both false alarms and low-power misses are statistically inevitable, rather than signs of pathology. Failure to accept this probabilistic viewpoint can contribute to a (false) feeling of crisis, and thence to unreasonable remedies. For example, in a passage that later became famous, Cohen (1962) wrote:

One can only speculate on the number of potentially fruitful lines of investigation which have been abandoned because Type II errors were made, a situation which is substantially remediable by using double or triple the original sample size.

A generation of researchers could be profitably employed in repeating interesting studies which originally used inadequate sample sizes. (p. 153)

*Profitably*, compared with what else they might have done? Ultimately, this thought yielded a generation of statisticians gainfully employed in calculating Type II error probabilities for biomedical studies. Cohen was a great and influential commentator, but in this case failed to accept fully the inevitable tradeoff among effect size, sample size, and probability of missing something worthwhile. How should one tell where to double or triple sample size? Studies are often judged as “interesting” only after the effect size has been estimated.

Cases where replication fails and ones where detection fails both have to be accepted as consequences of the laws of probability.

Valid replication often requires theoretical understanding of the phenomenon in question, which is attained only later. A failed replication study may differ from the original one by sampling from populations that differ on variables such as age, sex, experience, or culture that only later are seen to be relevant.

One should keep in mind that scientific progress depends in part on asking “the right” questions, ones that turn out to be fruitful. Such questions sometimes emerge from prior findings—even when the latter are smaller than first appeared—and sometimes from imaginative attempts to extend exciting theories into new domains. Both sources of progress are seen in modern psychological research.

## **A comment on the history of measurement foundations**

Measurement in psychology is founded on two distinct ideas, one drawn from geometry, the second from probability. Microeconomic utility theory, conjoint measurement, and Stevens’s (1936, 1966) scaling methods all belong to the geometric strand, while the probabilistic strand includes Fechner, signal detection theory, and Luce’s (1959) seminal monograph.

### ***Geometric strand***

Euclidean geometry is named for a famous axiom system, which systematized many (approximate) facts. Representation and uniqueness theorems (Cartesian analytic geometry) awaited the later development of number systems. In the 19th century, subtle

implicit assumptions were discovered and made explicit (e.g., Helmholtz, 1868; Hilbert, 1899). Most axiom systems took points, lines, incidence of points on lines, and congruence of intervals as the basic notions, but the 20th-century use of geometric models for dissimilarity (Shepard, 1962) gave rise to metric and analytic geometries with ordering of distance pairs as basic (*FOM*, Chapter 14).

It seemed natural to extend foundational work to physical quantities such as mass. Hölder (1901) noted the importance of an associative concatenation operation as basic; such an operation was later naively enshrined as the *sine qua non* for fundamental measurement of quantity. (Curiously, Newton's representation of force by 3-dimensional vectors long escaped foundational scrutiny except for a fragmentary 1-dimensional treatment by Euler. Only when vector representations for force and color were compared did it become clear that these shared a common axiomatic foundation in which color matching or static equilibrium are the basic notions; see *FOM*, Chapter 15.)

Stevens's (1936) earliest attempt to base loudness measurement on human "ratio" judgment was rejected by some, in part because there isn't a good way to concatenate two subjective loudnesses. This criticism is still reflected in Trendler's (2019) complaint that psychological measurement assumes "*test subjects are somehow capable of determining magnitudes of quantity of psychological attributes* [emphasis added]" (p. 114). Indeed, Stevens was never able to address such a criticism properly. It may have just seemed "obvious" to him that people could judge subjective ratios. These critiques can be addressed, however, by viewing the judgments as an *ordering of pairs* with respect to a property that satisfies axioms characterizing a ratio (see *FOM*, Chapter 4; Krantz, 1972; Shepard, 1981). Such a view is justified by the observed coherent properties of such judgments (Stevens, 1966). It construes the measurement foundations as similar to foundations of geometry.

Microeconomists recognized that additive utilities convert indifference curves to parallel straight lines. This was axiomatized by Debreu (1960). Luce and Tukey (1964) generalized this slightly under the label "additive conjoint measurement." At that time, this latter seemed to many a quite new idea. Researchers did not immediately recognize the close relationship of their article to the formulation by Debreu (which rested heavily on topological connectedness) or to earlier work on measurement based on tradeoffs between dimensions (Davidson, Suppes, & Siegel, 1957; Ramsey, 1931; Suppes & Winet, 1955).

A seemingly complex combination of associative concatenation operation (union of disjoint events) with additive utility was the key to the pioneering foundational work of Savage (1954). Chapters 5 and 8 of *FOM* sort out the roles of these two disparate elements.

### **Probabilistic strand**

We have already noted that statistical models for choice frequencies are fundamental to much of psychophysics. Fechner's approach to interval-scale measurement was formulated rigorously and stated clearly by Luce and Edwards (1958). Thurstone's (1927) scaling models and signal detection theory (D. M. Green & Swets, 1966) are closely related, but have additional valuable features. A culmination for this approach was the monograph, *Individual Choice Behavior* (Luce, 1959), whose profound influence on

psychological theory can hardly be overestimated. Luce dealt with choices among sensory inputs (psychophysics), with preference-based choices (decision making), and with changes of choice probability as a function of experience (learning). The monograph introduced ratio-scale measurement based on a testable property of choice probabilities (Luce's *Choice Axiom*) and applied this measurement principle to generate innovative theories in psychophysics, decision-making, and learning.

## Conclusion

We agree with Trendler (2019) that the effect of conjoint measurement on empirical practice was less than initially hoped for. In retrospect that hope was unrealistic. This does not diminish the impact that axiomatic analyses have on our understanding of the nature of measurement in the behavioral sciences.

Trendler's concerns go far beyond conjoint measurement to claim that interval and ratio scales do not exist in psychology, and as a result inferential statistics on behavioral data are meaningless, so replications fail and the field is stagnant. We strongly disagree with all points and linkages in this sequence.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## Note

1. We cite the three volumes of *Foundations of Measurement* (Krantz, Luce, Suppes, & Tversky, 1971/2007; Luce, Krantz, Suppes, & Tversky, 1989/2007; Suppes, Krantz, Luce, & Tversky, 1990/2007) simply as *FOM* throughout the comment.

## References

- Anderson, N. H. (1970). Functional measurement and psychophysical judgment. *Psychological Review*, 77, 153–170.
- Atkinson, R. C., & Shiffrin, R. M. (1969). Storage and retrieval processes in long-term memory. *Psychological Review*, 76, 179–193.
- Bartlett, F. C. (1932). *Remembering: An experimental and social study*. Cambridge, UK: Cambridge University Press.
- Cliff, N. (1992). Abstract measurement theory and the revolution that never happened. *Psychological Science*, 3, 186–190.
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, 65, 145–153.
- Cowan, N. (2008). What are the differences between long-term, short-term and working memory? *Progress in Brain Research*, 169, 323–338.
- Davidson, D., Suppes, P., & Siegel, S. (1957). *Decision making: An experimental approach*. Stanford, CA: Stanford University Press.

- Debreu, G. (1960). Topological methods in cardinal utility theory. In K. J. Arrow, P. Suppes, & S. Karlin (Eds.), *Mathematical methods in the social sciences, 1959* (pp. 16–26). Stanford, CA: Stanford University Press.
- Falmagne, J.-C., Iverson, G., & Marcovici, S. (1979). Binaural “loudness” summation: Probabilistic theory and data. *Psychological Review*, *86*, 25–43.
- Fechner, G. (1860). *Elemente der Psychophysik* [Elements of psychophysics]. Leipzig, Germany: von Breitkopfs Härtel.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.
- Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research*, *5*, 103–123.
- Helmholtz, H. von (1868). Über die Tatsachen die der Geometrie zu Grunde liegen [On the facts that underlie geometry]. *Göttingen Nachrichten*, *9*, 193–221.
- Higgins, E. T. (2005). Value from regulatory fit. *Current Directions in Psychological Science*, *14*, 209–213.
- Hilbert, D. (1899). *Grundlagen der Geometrie* [Foundations of geometry]. Stuttgart, Germany: Teubner.
- Hölder, O. (1901). Die Axiome der Quantität und die Lehre vom Mass [The axioms of quantity and the science of mass]. *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematische-Physische Klasse*, *53*, 1–64.
- Horner, M. D. (1990). Psychobiological evidence for the distinction between episodic and semantic memory. *Neuropsychology Review*, *1*, 281–321.
- Krantz, D. H. (1972). A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology*, *9*, 168–199.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (2007). *Foundations of measurement: Vol. I. Additive and polynomial representations*. New York, NY: Academic Press. (Original work published 1971)
- Kruskal, J. B. (1965). Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society, B*, *27*, 251–263.
- Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York, NY: Wiley.
- Luce, R. D., & Edwards, W. (1958). The derivation of subjective scales from just noticeable differences. *Psychological Review*, *65*, 227–237.
- Luce, R. D., Krantz, D. H., Suppes, P., & Tversky, A. (2007). *Foundations of measurement: Vol. II. Geometric, threshold, and probabilistic representations*. New York, NY: Academic Press. (Original work published 1989)
- Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, *1*, 1–27.
- Michell, J. (1999). *Measurement in psychology*. Cambridge, UK: Cambridge University Press.
- Moscovitch, M., Cabeza, R., Winocur, G., & Nadel, L. (2016). Episodic memory and beyond: The hippocampus and neocortex in transformation. In S. T. Fiske, D. L. Schacter, & S. E. Taylor (Eds.), *Annual review of psychology* (Vol. 67, pp. 105–134). Palo Alto, CA: Annual Reviews.
- Pais, A. (1986). *Inward bound: Of matter and forces in the physical world*. Oxford, UK: Oxford University Press.
- Ramsey, F. P. (1931). *Truth and probability*. In *The foundations of mathematics and other logical essays* (pp. 156–198). New York, NY: Harcourt Brace.
- Savage, L. M. (1954). *The foundations of statistics*. New York, NY: Wiley.
- Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *13*, 501–518.



- Schurgen, M. W. (2018). Visual memory, the long and short of it: A review of visual working memory and long-term memory. *Attention, Perception and Psychophysics*, *80*, 1035–1056.
- Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika*, *27*, 125–140.
- Shepard, R. N. (1981). Psychophysical relations and psychophysical scales: On the status of “direct” psychophysical measurement. *Journal of Mathematical Psychology*, *24*, 21–57.
- Slovic, P. (1995). The construction of preference. *American Psychologist*, *50*, 364–371.
- Stevens, S. S. (1936). A scale for the measurement of a psychological magnitude: Loudness. *Psychological Review*, *43*, 405–416.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*, 677–680.
- Stevens, S. S. (1966). Matching functions between loudness and ten other continua. *Perception & Psychophysics*, *1*, 439–446.
- Suppes, P., Krantz, D. H., Luce, R. D., & Tversky, A. (2007). *Foundations of measurement: Vol. III. Representation, axiomatization, and invariance*. New York, NY: Academic Press. (Original work published 1990)
- Suppes, P., & Winet, M. (1955). An axiomatization of utility based on the notion of utility differences. *Management Science*, *1*, 259–270.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*, 273–286.
- Trendler, G. (2019). Conjoint measurement undone. *Theory & Psychology*, *29*, 100–128. doi: 10.1177/0959354318788729
- Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*, 105–110.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453–458.
- Wallsten, T. S. (1976). Using conjoint-measurement models to investigate a theory about probabilistic information processing. *Journal of Mathematical Psychology*, *14*, 144–185.

## Author biographies

David H. Krantz was Professor of Psychology at the University of Michigan (1964 – 1979) and later Professor of Psychology and Statistics at Columbia University (1985 – 2015). Between these posts he was Member of the Technical Staff at AT&T Bell Laboratories. Since retirement in 2015 he has been Professor Emeritus at Columbia. His research at Michigan included foundations of measurement, vision research and psychophysics. At Bell Labs he concentrated on probabilistic reasoning. At Columbia his focus was decision making, especially the study of social goals and cooperation. He led the Experimental Psychology program for a time at Michigan. He was Chair of Statistics for a time at Columbia and was later a leader in Columbia’s Earth Institute and Center for Research on Environmental Decisions. In retirement his foci have been (i) foundations and practice of statistics, (ii) human cooperation and (iii) organizational goals and plans. His best-known publications are the three volumes of *Foundations of Measurement* (reprinted, Dover Publications, 2007).

Thomas S. Wallsten is Professor Emeritus and former department chair in the Department of Psychology at the University of Maryland. A cognitive and mathematical psychologist, he has published more than 100 articles and chapters primarily in the areas of probabilistic inference, judgment, choice, and communication under risk and uncertainty, topics that he still pursues in retirement. He is a former editor of the *Journal of Mathematical Psychology*, a former associate editor of *Psychometrika* and of the *Journal of Experimental Psychology: Learning, Memory, and Cognition*, and currently is on the editorial board of *Decision* and the editorial advisory board of the *Journal of Mathematical Psychology*.