Commentary

FURTHER COMMENTS ON THE "NONREVOLUTION" ARISING FROM AXIOMATIC MEASUREMENT THEORY

Louis Narens and R. Duncan Luce
Institute for Mathematical Behavioral Science, University of California, Irvine

Cliff's (1992) commentary on the failure of axiomatic measurement theory (AMT) to generate as much impact on cognitive psychology and psychometrics as he had once anticipated invites further commentary. For the most part, we do not disagree with his observations, but we believe that some amplification and clarification may be helpful. We attempt to establish three major points:

1. There are areas of psychology (e.g., decision making and psychophysics) in which AMT has had considerably more impact than Cliff acknowledges, and in these areas it assumes the form of theory, not scale, construction.
2. There are results of a new type (described in Luce, Krantz, Suppes, & Tversky, 1990; Narens, 1985), less well known than those of Krantz, Luce, Suppes, and Tversky (1971), about which Cliff makes no comment. These results should be of broad interest in psychology for two reasons: They provide a shelf of nonadditive representations that can be drawn upon along with the traditional additive and multiplicative ones, and they give better understanding about how to apply meaningfulness and invariance arguments.
3. The failure of measurement to "take" in cognition and psychometrics is related to a deep conceptual question concerning the relationship between statistics, as a way of describing randomness, and measurement, as a way of describing structure. The lack of an adequate theory for this relationship is, in reality, a weakness of both fields.

Our observations do not undercut Cliff's charge that a possibly major reason for the limited impact lies with the researchers themselves. There is no question that our published works tend to be mathematically accessible only to persons having some exposure to abstract algebra, geometry, and topology. Except for Roberts, whose 1979 book describes some of the major additive models and their applications, the field is still awaiting someone willing and able to write a suitable bridging work. We suspect this is a major reason why Cliff and others seem unaware of some of the important applications and developments of the past 10 years. We hope this challenge will soon be met.

Address correspondence to R. Duncan Luce, Social Science Tower, University of California, Irvine, CA 92717.

EXAMPLES OF USES OF AXIOMATIC MEASUREMENT THEORY

We discuss in this commentary a class of theories that have two primary intertwined goals: (1) to provide qualitative theories underlying quantitative models that relate several variables, and (2) to provide a theoretical foundation for measuring quantitatively the variables of a qualitative theory. The quantitative models characterize how the measurements of qualitative variables are linked, and the measurement axiomatizations explain how different qualitative experimental situations link together to capture the corresponding qualitative model. For each quantitative model, many measurement axiomatizations can be given, each corresponding to a different way of capturing the quantitative model in terms of experimentally observable primitives. Because of this, the insights embodied in formulating good axiomatic theories are often similar to those used in designing good experiments. Formulating such axiomatic, qualitative models places stringent requirements of clarity on the researcher.

Although the two goals of measurement theory are both theoretical, they yield powerful practical implications. The first is that the axioms characterizing the qualitative theory often suggest focused, qualitative experimental tests of the theory. These can pinpoint what is amiss, and such knowledge can provide a basis for modifying the theory. In contrast, if a quantitative model fails to give a good overall fit to a body of data, that failure often provides little direction about how to formulate a better theory other than by adding more free parameters.

The second implication is that such a theoretical foundation is useful not only in measuring variables, but also in determining their scale types. The impact of this observation is discussed in the next section.

Throughout, we cite examples illustrating applications of AMT. Although they are selected to demonstrate the diversity of problems AMT addresses, they are far from exhaustive.

Decision Theory

The dominant approach to studying individual decision making under risk and uncertainty began with von

Neumann and Morgenstern (1947) and Savage (1954). The primitive concept is the qualitative ordering generated by choices among sets of alternatives defined over chance events. The initial (normative) theories showed that if certain principles of consistency are met, the behavior is as if there is a utility function over consequences and a subjective probability measure over events such that choices among risky and uncertain alternatives involve selecting the one having the largest subjective expected utility. Although early empirical research tended to support this overall quantitative model as a good first-order approximation to human behavior, careful examination of the axioms revealed unambiguous difficulties (see, e.g., Kahneman & Tversky, 1979). Modified proposals led to increasingly better descriptive theories. Decision theory has been a textbook example of using axioms to localize the difficulty in a theory. The process of finding axiomatic failures and generating altered axiomatizations and new models for risky decisions has been repeated several times (Fishburn, 1982, 1988; Wakker, 1989).

Axiomatic measurement has been far more than a "revolution" in decision theory; it has essentially created the field. And it is a field with socially significant applications in business, operations research, and engineering.

Color Theory

The psychophysics of color matching is deeply embedded in the science of vision, and a variety of quantitative models have been developed to describe the "laws" of color matching. Exactly how these models are related to one another and what qualitative principles they encompass turned out to be a subtle matter that eluded researchers until Krantz systematically worked out axiomatic measurement theories for them (for a summary, see chap. 15 of Suppes, Krantz, Luce, & Tversky, 1989).

Magnitude Estimation

Procedures in which subjects provide verbal estimates of subjective intensities of stimuli are much used in psychology and other social and behavioral sciences. Despite the value of these procedures as a research tool, they have generated much controversy because a clear theoretical foundation was lacking. Without such a foundation, one cannot properly distinguish between valid and invalid uses of these methods. Krantz (1972), Luce (1990), Narens (1993), and Shepard (1981) have suggested somewhat different measurement-theoretic axiomatizations for magnitude estimation that allow one to determine the appropriateness of magnitude estimation for a wide range of situations.

NEW RESULTS

The past 12 years have witnessed a bit of a revolution within AMT itself. It has centered on scale type, invariance, and meaningfulness. The topic began with Stevens's (1946, 1951) classification of scale types—nominal, ordinal, interval, and ratio—and his contention that statistical propositions should be invariant under the transformations appropriate to the measures in question. His discussion suffered from there not being then a well-developed theory of what constitutes measurement, an explanation of what gives rise to his classification, and a systematic argument in defense of his invariance condition.

Scale Type and Admissible Statistics

Narens (1981a, 1981b) initiated a theory of scale types that is based on inherent properties of the qualitative measurement structure, and the relation of meaningfulness and invariance is now better (although not fully) understood in these terms. Within the class of homogeneous[1] structures with finitely many degrees of freedom and continuous real representations, Stevens's intuition that interval and ratio scales exhaust the possibilities was nearly correct. Between these two, however, lie other scale types, some of which exhibit underlying periodicities in the structure (Alper, 1987). In addition, the AMT analysis of meaningfulness has greatly muted the controversy about admissible statistics by clearly describing those kinds of inferences for which invariance arguments are appropriate and those for which such arguments are irrelevant. These ideas apply not just to statistics, but to a whole range of arguments found in the sciences, including dimensional analysis, geometry, and other parts of mathematics. Luce et al. (1990) summarize many of the topics, including a detailed discussion of admissible statistics (chap. 22).

Laws of Combination

These scale-type ideas play an increasing role in achieving suitable psychophysical laws (Falmagne, 1985) and in offering families of nonadditive representations. As an example of the latter, consider a binary operation $\bigcirc$ of combining two things exhibiting a common attribute

1. A structure is homogeneous if it is impossible, aside from their separate identity, to distinguish any two elements by differences in their structural properties. A structure with a maximum or a minimum element is not homogeneous because these elements are structurally unlike all others, which necessarily lie between them. Structures that admit ordinal, interval, or positive ratio representations are necessarily homogeneous.

to yield a new object also exhibiting that attribute. Often the order of the combination matters: $a \bigcirc b$ is not equivalent to $b \bigcirc a$. A familiar example of such a noncommutative operation is a weighted average with weights different from ½; other examples arise in the analysis of nonadditive conjoint structures. The simple fact of noncommutativity means that the operation $\bigcirc$ cannot possibly be represented by numerical addition, +. What are the options? For homogeneous situations, they are simply ratio scale representations of $\bigcirc$ by a numerical operation $\oplus$ that has the form:

$$x \oplus y = y\mathrm{f}(x/y),$$

where f is strictly increasing and $\mathrm{f}(z)/z$ is strictly decreasing. The task becomes one of either characterizing f theoretically or estimating it empirically. Similar simple representations exist for two-factor structures (Luce et al., 1990, pp. 180–184).

Merging of Ratings

In contests of various sorts, in selecting among candidates for a position, and in other settings, judges' ratings of a set of alternatives must be merged to form a consensus rating. The arithmetic mean of the ratings is commonly used. It has been shown that if the ratings of individual judges are assumed to form ratio scales, then arithmetic means are valid, in the sense of yielding the same overall ranking independent of the particular units selected for each of the individual scales, only if a very controversial theoretical assumption holds, namely, that the judges' subjective intensities underlying the ratings lie on a common subjective scale. Such an assumption of interpersonal comparability is exceedingly suspect and very difficult to defend (Narens & Luce, 1983). Abandoning that assumption, one can show that merging based on geometric means is valid and that any other valid method is equivalent to it (Aczél & Roberts, 1989).

RANDOMNESS AND STRUCTURE

Put somewhat baldly, statistics focuses mostly on randomness, largely taking structure among variables for granted; AMT focuses almost exclusively on structure, largely ignoring randomness. These are two partial facets of a single problem: to uncover structure among variables in the presence of inherent randomness. Roughly, statistics presumes numerical observations and the form of the structure (linear, log-linear, multinomial, euclidean, factor analytic, etc.) underlying them, and it attends to the numerical representation of randomness within that framework. Typically, statistical trouble arises when one considers the possibility of general monotonic transformations of aspects of the data in order to explore structure more fully.[2] For example, such monotonic transformations are a major feature of the psychometric technique of nonmetric multidimensional scaling, but its statistical treatment is rather casual. (Axiomatizations of nonmetric multidimensional scaling are at this time only partially satisfactory—e.g., see Suppes et al., 1989, chap. 14).

By contrast, AMT begins with ordinal and relational data and postulates properties (i.e., axioms) that the data are believed to satisfy in one or more empirical interpretations. Its goal is to determine whether the data can be represented by some numerical structure, such as the ordered structure of the additive, positive real numbers. This approach simply does not acknowledge randomness in its formal description of the data from which structure is to be inferred. True, a few measurement studies try to accommodate the numerical representation to the variability of the data, but such treatments are not really a part of the basic description of qualitative structure. For example, one approach takes the probabilities of choices from sets of alternatives as primitive, but it has proved very difficult to incorporate more than ordinal structure into this framework (Suppes et al., 1989, chap. 17).

Neither camp has any very clear idea about how to formulate the concept of randomness in a nonnumerical way. In particular, we simply do not know how to talk about randomness at a level involving only qualitative ordinal observations together with other qualitative relationships, such as combining two stimuli to form a third. Yet this is the level at which measurement theorists believe structural questions must be understood. We need a theory for the simultaneous qualitative description of both structure and randomness that leads to a representation not into numbers, but into families of random variables. Their expected values would form the domain of a numerical structure describing qualitative structural relationships, much like the representations of current AMT, and their distributions would constitute the basis of the corresponding statistical theory.

Although we believe that the existing incompleteness of statistical and measurement approaches largely accounts for the limited impact of AMT in psychometrics, it only partially accounts for why AMT has been ignored in many areas of cognitive psychology. Often the variables studied by cognitive psychologists are or are treated as categorical, usually binary. The question typically posed is whether a change in one variable affects another. It is rare for the focus to be on the structural relations among the several independent and dependent variables. Increasingly, structural issues are being raised

2. Nonparametric methods avoid structural commitments and do not focus on structure.

about categorical variables. For example, Riefer and Batchelder (1988) have strongly questioned the suitability for much cognitive data of the usual linear models and analysis-of-variance or regression approaches. Instead, they have proposed, and analyzed statistically, an alternative structural formulation known as *multinomial modeling*. As cognitive research attends more to structural issues, it will have greater reason to be interested in measurement theory, although it surely will be badly frustrated by the latter's current failure to incorporate randomness explicitly.

CONCLUSIONS

Our aim has been threefold: First, we sought to establish that the applications of AMT are considerably wider and hold more potential than recognized in the critique by Cliff (1992). Second, we cited some of the newer, less well known results that not only clarify deep issues raised by psychologists, but also offer interesting ways to model functional relations among variables. And third, we proposed that AMT's limited impact in psychometrics arises, in part, from the lack of a qualitative theory of randomness, and AMT's limited impact in cognitive psychology arises, in part, because of the latter's limited concern about structural relations among variables.

Acknowledgments—We thank J.-C. Falmagne and A.A.J. Marley for comments on an earlier draft of this material.

REFERENCES

Aczél, J., & Roberts, F.S. (1989). On the possible merging functions. *Mathematical Social Science, 17,* 205–243.

Alper, T.M. (1987). A classification of all order-preserving homeomorphism groups of the reals that satisfy finite uniqueness. *Journal of Mathematical Psychology, 31,* 135–154.

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

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

Fishburn, P.C. (1982). *The foundations of expected utility*. Dordrecht, The Netherlands: Reidel.

Fishburn, P.C. (1988). *Nonlinear preference and utility theory*. Baltimore, MD: Johns Hopkins Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291.

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. (1971). *Foundations of measurement: Vol. I. Additive and polynomial representations*. New York: Academic Press.

Luce, R.D. (1990). "On the possible psychophysical laws" revisited: Remarks on cross-modal matching. *Psychological Review, 97,* 66–77.

Luce, R.D., Krantz, D.H., Suppes, P., & Tversky, A. (1990). *Foundations of measurement: Vol. III. Representation, axiomatization, and invariance*. New York: Academic Press.

Narens, L. (1981a). A general theory of ratio scalability with remarks about the measurement-theoretic concept of meaningfulness. *Theory and Decision, 13,* 1–70.

Narens, L. (1981b). On the scales of measurement. *Journal of Mathematical Psychology, 24,* 249–275.

Narens, L. (1985). *Abstract measurement theory*. Cambridge, MA: MIT Press.

Narens, L. (1993). *A theory of magnitude estimation*. Manuscript submitted for publication.

Narens, L., & Luce, R.D. (1983). How we may have been misled into believing in the interpersonal comparability of utility. *Theory and Decision, 15,* 247–260.

Riefer, D.M., & Batchelder, W.H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review, 95,* 318–339.

Roberts, F.S. (1979). *Measurement theory with applications to decision making, utility, and the social sciences*. Reading, MA: Addison-Wesley.

Savage, L.J. (1954). *The foundations of statistics*. New York: Wiley.

Shepard, R.N. (1981). Psychological relations and psychophysical scales: On the status of "direct" psychophysical measurement. *Journal of Mathematical Psychology, 24,* 21–57.

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

Stevens, S.S. (1951). Mathematics, measurement and psychophysics. In S.S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1–49). New York: Wiley.

Suppes, P., Krantz, D.H., Luce, R.D., & Tversky, A. (1989). *Foundations of measurement: Vol. II. Geometrical, threshold, and probabilistic representations*. New York: Academic Press.

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton, NJ: Princeton University Press.

Wakker, P.P. (1989). *Additive representations of preferences: A new foundation of decision analysis*. Dordrecht, The Netherlands: Kluwer Academic Publishers.