Separating the Wheat from the Chaff: A Review of Some Commentaries Concerning the Conceptual and Empirical Status of Kinetic Output

J. Powell

California University of Pennsylvania

How odd it is that anyone should not see that all observation must be for or against some view if it is to be of any service.

—Charles Darwin

I would like to thank the previous editor of *The Behavior Analyst*, Sigrid Glenn, for moving our manuscript through the review and acceptance stages and the current editor, Sam Deitz, for completing the editorial process. It was Professor Deitz's idea to solicit commentaries and a rebuttal, an idea that I greeted with considerable enthusiasm—that, of course, was prior to receiving any of the commentaries. Actually, I have profited substantially from the ideas put forth by a distinguished group of behavioral researchers. As the title of this rebuttal implies, I found many of the comments insightful and provocative, while others seemed to me to be wide of the mark, in some cases substantially so. I will attempt to elaborate on both categories.

Marr seems to me to have things just about right. He makes an important point when he argues that we employed a dimensional analysis within a framework that lacks a referent dimensional system. That is, in the behavioral sciences there is not now nor does it seem likely that there will soon be a dimensional system analogous to [MLT]. Lacking a dimensional system we were, as Marr suggests, free to play with some standard behavioral quantities to see if something useful might emerge. The discovery process was less random than it might first appear,

though, since a major restraint served to restrict the number of possible outcomes: Any product of the analysis had to exhibit dimensional integrity. One reviewer of the manuscript stated that there were a number of different possible combinations of N , D and T . Why not consider them all? The answer is that one ends up with a variety of noninterpretable entities. For example, the product of $N \times D \times T$ yields the quantity $r \text{ min}^2$, a result that appears to be nonsensical in the extreme. The product of $N/(D \times T)$ does have dimensional integrity for it yields a notation, r/min^2 which is dimensionally equivalent to response acceleration. That this product has nothing in common with response acceleration (changes in rate over time) other than a shared notation was obvious, and we therefore dismissed it as a possibility. The only permissible grouping to emerge was the one we proposed and evaluated.

Buckingham's theorem states that, if there are m physical variables expressed by n fundamental quantities, then $m - n$ dimensionless groups may be formed. Marr is thus correct in stating that our formulation suggests three dimensionless groups only one of which is accounted for, and he wonders what these missing groups might be. So do I. The search for these "hidden" groups must begin by considering what other quantitative properties are intrinsic to behavior that may contribute to an observed output. A prime candidate is some expression that reflects the energy expended by the organism. We suggested that relative weight (W_r), the individual's weight divided by some standard weight, might serve as a suitable surrogate expression since energy expenditure is proportional to body size. The output equation would then assume the form $z \log\{[(N \times D)/T] [W_r]\}$.

Author Note: Reprints may be obtained from the author, Honors Program, Box 100, California University of Pennsylvania, California, PA 15419.

If the explicit referencing of another dimensionless group accounted for additional variability, then the variable we called z would show increased stability. I would not want unnecessarily to antagonize the behavioral community, but it should be pointed out that the process described here and in the original paper would remain unchanged if we denied physical status to z , an assumption that would eliminate the requirement of specifying the number of dimensionless groups that needed to be determined.

Order may be discovered through observation or experimentation. Marr notes that for behavior analysis the major focus is on the latter process whereby one seeks to establish systematic relationships between environmental variables and some measure(s) of behavior. The primary issue according to Marr, therefore, is not the stability of the measure but the ability of the measure to enter into effective functional relationships. Is there any evidence that K_o would yield effective functional relationships? My answer is yes, for if we confine the analysis to basic research findings I am unaware of any existing relationship based on rate that would not be confirmed if K_o were substituted for rate. This conclusion is based on the inextricable relationship that exists between frequency and duration in basic research studies. Because each response, e.g., a key peck, is of some brief duration, the two quantities will always move in the same direction. For example, increases in response rate will always be matched by corresponding increases in total response duration; it follows that K_o would also increase. If this view is correct it would be accurate to say, for instance, that as the density of reinforcement increases, K_o increases, or as the intensity of the punisher increases, K_o decreases. It is comforting to recognize that our formulation can accommodate existing findings, a point relevant to some issues raised in other of the commentaries.

I want to address briefly an example offered by Marr, for it helps to explain why one would be interested in exploring alternative quantitative expressions. Marr

notes that our ability to predict the course of behavior under a fairly simple circumstance—resistance to extinction—is woefully inadequate, and he attributes this failing, at least partly, to our incomplete understanding of the dynamics of behavior. Exactly so. I regret that we were not clever enough to state the case as clearly as Marr did. It is not incumbent on an alternative formulation, as several of the reviewers have suggested, to lead to new relationships: An incremental advancement in our understanding of existing relationships would be enough. Are there circumstances where our formulation or some expanded variation of it would lead to more predictable relationships? I do not dismiss it as a possibility for it is a question that can only be answered empirically.

Marr raises several additional points that I will briefly comment on. The terms, stereotypic and attending, used to categorize the observed behaviors are generic terms that do not reflect the diversity of the various topographies exhibited by individual subjects. Summing the results across subjects would not seem very useful under these circumstances. It may well be that the mixed results are at least partially attributable to this lack of uniformity in the behaviors observed and recorded. It would be inappropriate to introduce additional findings here, but I have previously in a public forum (Powell, 1986) presented the frequency distributions for K_o for the data reported in our study. It is fair to say that these distributions approximate normal distributions, an initially encouraging finding because it speaks to the issue of the behavior of K_o .

Greenspoon and Olson offer the interpretation that we reduced variability through a mathematical manipulation. Not so. What we did was propose an alternative quantitative expression, evaluated the expression across behaviors and subjects and then examined the variability of the results. Let us not forget that there were instances where rate and duration exhibited less variability than K_o , a result that presumably could lead Greenspoon and Olson to conclude that

via a mathematical manipulation we produced an expression that increased variability. Although ambiguous results are not often welcome, in this case they were since they demonstrated that our formulation did not demand that K_0 reduce variability, a point seemingly overlooked by Greenspoon and Olson. A related issue raised in their commentary is that the behavior per se wasn't more stable, only the transformed data were. Presumably this means they consider rate and duration measures to represent independent and immutable behavioral "truths," not a position I am comfortable with. To propose as we did that frequency and duration can be members of an intersecting set is to assume that some sort of interaction exists between the members.

Greenspoon and Olson point out that the behavior analytic approach encourages the search for and identification of sources of variation. Once identified, the variables contributing to the instability may be held constant, eliminated or, in the more interesting case, lead to the discovery of new independent variables. Agreed. What these authors apparently find difficult to accept, however, is the notion that the way we conceptualize and subsequently describe behavior processes may be a source of the observed variability. Or, more to the point, in some instances the multifaceted dynamic structure of behavior may not be adequately represented by a single metric, be it rate or duration. To suggest as we did that output may be integrated at a level other than rate or duration is to make the same point. And, finally, the search for order is not limited to examining the relationship between independent and dependent variables. Any descriptive science, e.g., geology, could be used as the model for the approach we took, which is to assume that the natural phenomenon to be investigated—in this case behavioral output—is inherently orderly. Guided by theory or less formal considerations, e.g., educated guesses, observations are employed in an attempt to uncover that order.

Timberlake and Dougan argue that we

provided no conceptual, theoretical or mathematical assumptions to justify our integrated expression. I thought we had, but let me briefly restate the case. Collapsing the two metrics, frequency and duration, into a unified expression is possible because of the dimensionless status of a frequency measure of behavior. Consider the difference between the two statements, 10 responses and 10 m, the latter specifies both an arbitrary standard, the meter, and a dimension, length. The former simply names the thing being counted as in, for instance, 10 chickens. The term frequency as used in the behavioral sciences means measurement by enumeration, whereas in the physical sciences it is a quantity derived from a fundamental dimension, time. This brief discussion is meant to demonstrate that attention to the dimensional status of common behavioral measures guided creation of the expression, not arbitrarily choosing one we found pleasing as the authors suggest. To repeat, any grouping other than the one we proposed may be dismissed for logical or conceptual reasons.

Timberlake and Dougan acknowledge that rate and duration are not well integrated empirically or conceptually. They understate the case. An understanding of the relationship between the two metrics is virtually nonexistent. How could it be otherwise, since so few studies employ both measures? The studies we cited were informative only to the extent that they demonstrated no consistent relationship between the two. Timberlake and Dougan unlike Johnston recognize the problematic issue this observed noncorrespondence raises. They outline a number of empirical and theoretical approaches that might be taken to examine the relationship between rate and duration, including exploring different ways of combining them: Our paper describes one preliminary means of doing so. Timberlake and Dougan argue that there is no a priori basis for supposing that a combined measure will be superior to a singular measure. I would only point out that there is also no a priori basis for assuming that it would not. If future em-

pirical work demonstrates that our expression or some evolved variant of it permits more effective prediction about behavioral phenomena, it will survive; if not, it will be abandoned. That is exactly as it should be.

A number of examples provided by Timberlake and Dougan purport to show that the use of K_o would produce results contrary either to expectations or already confirmed findings. In several of these examples it is simply unclear to me how the use of K_o would produce the results they say it would. Surely it is an inadequate test to find that our expression does not agree with data fabricated by the authors. My major disagreement, however, is based on the proposition advanced earlier, which is that K_o would yield the same relationship as rate in investigations that employ an arbitrary response. I was unnecessarily equivocal earlier when I suggested that, if the reasoning leading to this assertion were correct, it would be true. The reasoning was correct. K_o could be substituted for rate and leave many important functional relationships unchanged. The ability of our derived expression to accommodate existing findings is a fortunate by-product, for it played no role in the development of K_o . It might be taken as some slight evidence, though, that we are on the right track.

For illustrative purposes consider two sets of numbers, each set containing three pairs: 1,3; 2,2; 3,1 and 1,30; 2,20 and 3,10. The correlation between the pairs in each set is -1.00 , but the products of the pairs in the two sets are substantially different. If we assume the first member of each pair represents frequency and the second duration, it is easy to see that the product of the pairs (equivalent to K_o) has little or no relationship to their correlation. Timberlake and Dougan are mistaken when they speculate that a negative correlation between frequency and duration may be responsible for those instances where K_o variability was less. On the contrary, there is nothing to prevent the greatest variability in K_o from occurring when the correlation between frequency and duration is -1.00 . A corollary is that large differences in the value

of K_o can be produced from identical correlations. In conclusion, it is the magnitude of the pairs not their proportionality that influences K_o . As an aside, only three of the 20 unreported correlations in our study were negative.

I will simply list some items in Johnston's paper that I view as irrelevant, trivial or wrong concerning the question we addressed: response class, definition of a response, the contingencies involved in timed vs. untimed tests, the abbreviation used for our expression, the use of the coefficient of variation, and what "level of analysis" signifies.

It is difficult to think of a more exciting "scientific mystery story" than the long search for harmony in planetary movement. The geocentric model of Ptolemy with its circles, epicycles and offsets was supplanted by the heliocentric model of Copernicus with its circles, epicycles and offsets. And, finally, on to Kepler and his description of the elliptical orbit of the planets. This most elegant example of the search for and discovery of order was accomplished without manipulating a single variable—no independent variable, anywhere. Johnston's myopic view that order is relevant only within the context of an independent-dependent variable relationship is, as the above example is meant to show, patently false. Our effort to discover order through observation was not only legitimate, but also the preferred starting place. If the kinetic output expression had routinely demonstrated greater variability than rate or duration, it is unlikely one would want to use it to examine other behavior-environment interactions. Readers may wish to struggle on their own with the cryptic comment by Johnston that the evidence of reduced variability is different from reduced variability. Also, the decision by Johnston and his colleagues to discontinue measuring duration in their studies on rumination is mildly interesting, but not germane to the question we addressed. To state the obvious, the interaction between frequency and duration requires that they both be available for analysis.

Data acquisition methods need to be congruent with the analysis to be per-

formed. If rate and duration per session are to be reported, then the knowledge of the temporal location of a response in the first instance or the duration of each response in the second instance are unnecessary. Counts as provided by a cumulative counter and running-time meter, for example, will suffice. To record onsets and the number of times the behavior is recorded as present yields precisely what we indicated, which is rate and duration measures, respectively.

We may measure the force required to elongate an iron bar, or place an object on a scale and read its value in grams, or closer to home, measure the duration of social contact between two organisms. If we were to ask what is being measured in each of these instances, we would accept that it is elasticity, mass, and parental investment, respectively. This example is meant to demonstrate that a measure is always a measure of some underlying quality, a quality that experience has brought to our attention, that we elevate to the status of interesting, and, subsequently, make the object of further inquiry. Marr as well as Timberlake and Dougan are clearly aware of the importance of the conceptual basis of behavioral measurement; Johnston apparently is not. Habit strength, reflex strength, response strength, potential, tendencies,

etc., are some of the terms that have been used to describe the underlying quality we seek to elucidate. A reviewer suggested that if the concept of response strength presented a problem we ought to throw it out. If we were to do so, it would be at the expense of abandoning the scientific goal of prediction. We do not predict rate and/or duration values. We predict a relationship, conduct a suitable investigation, and then examine our measure(s) to see if the prediction is confirmed, denied, or left unanswered. To repeat what we said in our paper, the conceptual basis for rate and duration measurement is the same—an assessment of the strength of the response. What little evidence is available demonstrates there are instances where the two metrics yield equivocal interpretations about this underlying quality. Johnston sees no problematic issue here; I do. It suggests to me that the conceptual basis for our measurement, or the measures we employ, or some combination of the two factors may be inadequate under some circumstances.

REFERENCE

- Powell, J. (1986, May). *Behavioral output: A conceptual and empirical analysis*. Paper presented at the meeting of the Association for Behavior Analysis, Milwaukee, WI.