

Individual differences in perceived exertion assessed by two new methods

LAWRENCE E. MARKS

John B. Pierce Foundation Laboratory and Yale University, New Haven, Connecticut

and

GUNNAR BORG and GUNILLA LJUNGGREN

University of Stockholm, Stockholm, Sweden

Each of 30 male subjects judged, in a single session, the loudness of a 1000-Hz tone and the exertion perceived while pedaling a bicycle. Two psychophysical methods were used—one employing a combined category-ratio scale whose upper limit was defined as “maximum sensation” and the other a freer magnitude-estimation scale having no verbal labels. Both methods yielded data consistent with power functions, although the combined category-ratio scale gave slightly smaller exponents. The category-ratio estimates provided a measure of individual differences in perceived exertion: At any work level, the differences across subjects in judgment correlated with differences in heart rate (a physiological indicant of strain); this result is consistent with Borg’s hypothesis that in dynamic work, maximal sensation is at least roughly equivalent across subjects. When the magnitude and the category-ratio estimates were converted to equivalent loudness (Stevens and Marks’s method of magnitude matching), the derived loudness values also correlated with heart rate: This outcome provides evidence for the utility of the cross-modal procedure and provides further evidence consistent with Borg’s model of perceived exertion.

The goal of this study was to evaluate individual differences in the perception of exertion during dynamic work. That people can and do vary widely in their capacity to perform physical work goes without saying: Some individuals can achieve work levels that greatly exceed the capacity of others. What is less clearly established is the relationship, if any, between a given individual’s work capacity (e.g., the maximal physical level attainable) and his or her perception of exertion.

To compare the perceived exertion, or any other subjective magnitude, across individuals is not a simple matter, to say the least. Many popular scaling methods, such as magnitude estimation, provide no means to compare the suprathreshold judgments of one subject with those of another.

Indeed, some have questioned the very possibility of comparing sensation magnitudes across individuals. To be sure, it is impossible, as several have noted (e.g., Borg, 1961; Teghtsoonian, Teghtsoonian, & Karlsson, 1981), to make interindividual comparisons with psychophysics’s ultimate direct operation, namely, matching: An observer can directly compare sensations of his or

her own, but not his or her sensations to someone else’s. Nevertheless, we argue that statements comparing sensations across individuals can be meaningful: Most, if not all, philosophically/scientifically coherent positions vis-à-vis the relation of brain/body to mind (whether monistic or dualistic; reductive, epiphenomenal, or parallel) acknowledge a one-to-one or many-to-one relation between physiological and mental states. The search for neural correlates to perception is predicated on just such a view. The alternative would be to argue that the identical physiological state can be associated in different individuals with different mental states; such a position gives an ontological status to mental events that, we suspect, most scientists would find anathema.

Let us grant, then, that it is not meaningless to compare individuals with respect to sensation magnitude. Still, because empirical tests cannot be made *directly*—even were we able to measure brain states completely, we still would not, according to most philosophical positions, be comparing *perceptions* themselves—such comparisons are likely to arise, if at all, from theoretical considerations, leading to indirect tests. As with other theoretical constructs, we ask, first, “Is there a coherent theoretical framework to use in making interindividual comparisons?” and second, “Can statements about these comparisons yield testable predictions?”

To this end, Borg (1961, 1962, Note 1) suggested a simple principle—namely that individuals will experience the same degree of subjective exertion when they per-

This research was supported by a grant to G.B. from the Swedish Council for Research on the Humanities and Social Sciences, F407/81. It was conducted while L.E.M. was a visiting scientist at the University of Stockholm. L. E. Marks’s mailing address is: John B. Pierce Foundation Laboratory, 290 Congress Avenue, New Haven, Connecticut 06519. G. Borg and G. Ljunggren are in the Department of Psychology at the University of Stockholm.

form dynamic work at their respective maxima. If this is so, then all individuals are "calibrated," in a sense, in accord with their relative maxima: The maximum sensation of exertion is the same for everyone, although different people need different work levels to achieve this maximum. If, moreover, the form of the psychophysical (power) function relating perceived exertion to the physical level is the same in all individuals, then a second principle follows—namely that all people will experience the same degree of exertion when they perform at the same *percentage* of their respective physical maxima.

Borg's proposal that maximal sensation is constant is displayed graphically in Figure 1, which shows theoretical psychophysical curves for two individuals. Both curves express perceived exertion as a 1.6 power of the stimulus level (Borg, 1961), but with different scale factors, reflecting the fact that Subject 1 has a greater maximal physical capacity than does Subject 2. The hypothesis is that when the stimulus is calculated in terms of maximum capacity, perceived exertion for both subjects will be the same function of the percentage of maximum.

Neither principle—that perceived maximum is uniform; that the psychophysical function is uniform—is likely to be exactly true. Some variability undoubtedly exists, even in normal, healthy individuals. Variation in maximum alone would disturb interindividual equivalences throughout the dynamic range; variations in psychophysical function alone probably would disturb equivalences more at lower submaximal levels. Despite such variability, we hypothesize that maximum sensation is uniform enough to have not only testable but sizable consequences.

The utility of the "idealized" model comes in part from the predictions it makes in comparison with the

main alternative—namely, that maximal sensation does vary, in direct relation to strength (maximal *physical* output). [A third alternative—namely that both perceived maximal exertion and physical maximal capacity vary considerably, with no correlation between them, reduces to Borg's hypothesis of constant (strength-independent) maximum sensation, to which is added considerable variability; acknowledging that some variability exists, alternatives 1 and 3 differ only quantitatively.] The two main alternatives, constant maximum sensation and strength-related maximum sensation, can yield different predictions about performance, as described below.

The principle of constant maximal sensation has been embodied in the use of rating-scale judgments to evaluate perceived exertion. Although people may require different physical levels to attain maximal sensation, if a rating scale is anchored at the top by the category "maximal," then, following the "idealized model," judgments made with respect to maximum—say, as a percentage—should make scale values comparable across subjects. Moreover, continuing in this vein, categories such as "very strong," "slightly strong," and so on, should represent equivalent locations with respect to maximum sensation, despite the fact that different people require different work levels to produce a given subjective category. At least, this should be so given (1) that everyone has a similarly functioning physiological system, (2) that the verbal categories have similar meanings to everyone, and (3) that everyone has experienced similar ranges and distributions of work levels, so that they bring to the task a similar "adaptation level" (Helson, 1964). To the extent that these are so, then the use of an appropriate set of categories will serve to "normalize" psychophysical functions across individuals. Naturally, there will be some variability in the component processes. Still, it is reasonable to ask whether the hypothesized constancy of maximal sensation can show itself as a first-order approximation.

Reasoning along these lines, Borg (Note 1) obtained two types of judgment of perceived exertion during bicycling. One involved judgments as percentage of maximum. The other involved judgments on a particular category scale (called the ratings of perceived exertion, or RPE, scale, Borg, 1970), which is a linear function of power output; on the RPE scale, numerical values range from 6 to 20, to match the variation in heart rate from about 60 beats/min at rest to about 200 beats/min at maximum (7 on the scale is designated as "very very light" and 19 as "very very hard"). Judgments on the two scales were similar; in both cases the judgments made to any given stimulus level correlated significantly with heart rate, a physiological measure that varies directly (linearly) with power output.

Using heart rate as a measure of strain across individuals, then, the significant correlations provide indirect evidence for the hypothesis of uniform maximum sensation. The main alternative, that maxima vary with strength, would predict correlations of zero. It is notable

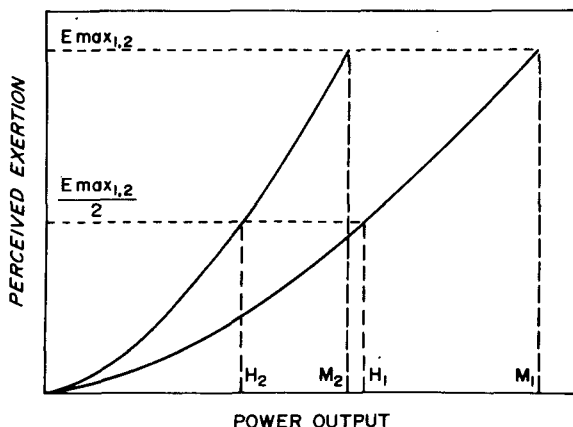


Figure 1. Theoretical psychophysical functions relating perceived exertion to power output during bicycling. According to Borg's (1961, Note 1) principle, subjects with different maximal capacities (E_{\max}) should experience the same levels of perceived exertion at their respective maxima and at constant percentages of their maxima.

that the size of the correlations decreased as work level decreased, a finding that would be consonant (1) with heart rate being more variable at low levels, and (2) with some individual differences in the underlying psychophysical functions.

The present study sought to assess individual judgments of exertion during bicycle ergometry and to compare the individual assessments obtained by means of two relatively new procedures that have been proposed as ways to make comparisons across individuals. One involves the use of a combined category-ratio scale, like that presented by Borg (Note 1). The RPE scale used earlier (Borg, Note 1), being a category scale, relates to power output by an exponent of 1.0, rather than the 1.6 that characterizes magnitude estimation. In the present experiment, the combined scale may serve to define sensations in terms of categories like RPE (with the upper limit defined as "maximal"), yet at the same time to maintain the ratio-like properties of magnitude-estimation scales and give an exponent near 1.6. The second new procedure, the method of magnitude matching, asks subjects to judge in a single session the sensation magnitudes produced by stimuli that activate two or more perceptual modalities (J. C. Stevens & Marks, 1980). This procedure permits one to compute individual cross-modality matching functions. Under the assumption that subjects' judgments of the intensities of sensations of different modalities can be represented on a single scale, and under the assumption that the psychophysical function on one of the modalities is the same in all subjects, the cross-modality matching functions make it possible to compare individuals on the other modality, and thus to assess individual differences.

In the present study, subjects judged both loudness and perceived exertion on the same scale, in one case a magnitude-estimation scale and in the other a category-ratio scale defined in terms of "maximum" sensation. The category-ratio scale itself may provide one means to compare individuals. In addition, assuming that the loudness function is the same in people with normal hearing, it should also be possible to assess individual functions for perceived exertion in terms of equivalent loudness. Again, it is unlikely that this will be completely correct. Nevertheless, the assumption that the loudness function is wholly constant makes it possible to examine implications (with respect to correlations with heart rate) in the limiting case. To the extent that individuals do differ in their loudness functions, independently of any differences in perceived exertion, the expected correlations will be reduced in size.

METHOD

The two scaling methods were used within a single test session for each subject. We tested all subjects using the same sequence of scaling tasks—first, magnitude estimation and, second, category-ratio scaling. We did this because subjects have a greater tendency to be "biased" by categorical procedures on a subse-

quent magnitude-estimation task than by magnitude estimation on a subsequent categorical task: After being given a fixed response scale in one task, subjects tend to use the same scale when subsequently given instructions to do magnitude estimation; but subjects can readily shift in the other direction, from the relatively free magnitude estimation to the fixed category scale. Given this asymmetry of bias, it was deemed better to use magnitude estimation first with all subjects, lest the results of the two parts of the session turn out to be too much alike.

Stimuli

For perceived exertion, the stimuli were produced by having the subject pedal on a calibrated bicycle ergometer, whose power level could be set to fixed values. Subjects pedaled at a rate of about 60 revolutions/min; this was done by instructing the subject to pedal at the rate needed for the visible needle on a gauge to reach an appropriate mark. However, the system is constructed to maintain constant power output even if the pedaling rate varies (over the range of 45-70 rpm).

In the magnitude-estimation part of the session, we used two power levels—65 and 163 W (400 and 1,000 $\text{kpm}\cdot\text{min}^{-1}$)—and for category-ratio scaling, six levels—33, 65, 98, 131, 163, and 196 W (200-1,200 $\text{kpm}\cdot\text{min}^{-1}$). The subject's heart rate was monitored with ECG electrodes whose output could be read on an analogue meter. Heart rate was recorded at the beginning of each part of the session and just prior to those times when the subject made his judgments of perceived exertion.

For loudness, the stimuli were binaural pulses of a 1000-Hz tone. Signals were generated by a Philips oscillator (4-msec rise), amplified by a Sentec amplifier, timed (1 sec) by a Dekadist timer, and fed to calibrated Maico headphones mounted in Grason-Stadler cushions. The levels were set by adjusting a decibel attenuator to produce 12 SPLs, from 40 to 95 dB in 5-dB steps.

Procedure

The subject was seated on the bicycle ergometer (Siemens-Elema AB) throughout the session, except for the period between the two parts when he was free to move about. In the first part of the session, the method of magnitude estimation was used. The subject was told that he would be judging the loudness of tones and the perceived exertion of cycling; he was to assign to the first stimulus (a tone) whatever number he deemed appropriate to stand for its subjective magnitude (loudness); then to succeeding stimuli, tones and work levels alike, he was to assign other numbers in proportion to subjective magnitude. That is, all sensations were to be judged on the same numerical scale, although the subject chose the actual scale for himself. A special point was made that the subject should use a single scale, such that the same number would be assigned to a tone and to a work level if the loudness equaled the perceived exertion. This extension of magnitude estimation to two modalities, making possible cross-modal comparisons, is the method of magnitude matching (J. C. Stevens & Marks, 1980).

This first part of the experiment went as follows. First, 6 of the 12 SPLs were presented, one at a time, for judgment. Then the first work level (power output = 65 W) was given, and the perceived exertion judged during the last 15 sec of the 4-min bout of cycling (heart rate was recorded just prior to the judgment). After about a 30-sec rest, 12 SPLs were presented and their loudness judged, after which the second work level (power output = 163 W) was given and the exertion judged. Finally, 6 more SPLs were presented and judged. Thus, in the first part, each work level was judged once, each tone twice. The same ascending order of work levels was given to all subjects because of the development of fatigue, which is not fully dissipated even after a rest of a couple of minutes; were the reverse order used, this fatigue would differentially affect the judgment of the lower work level to a much greater extent. Order of SPLs, however, was random.

After the first part was completed, a 15-min rest period ensued, during which the subject was free to move about. Then the second part of the experiment began.

In the second part of the session, the subject used a response scale that combined numbers (set out in a ratio fashion) with verbal categories. As can be seen in Figure 2, the scale is bounded at the bottom by the verbal label "nothing at all" and by the number zero. At the top, the scale is bounded by the verbal label "maximal"; although no number is designated for this label, extrapolation from lower numbers suggests a value of about 15. The other verbal labels comprise the terms "extremely weak," "very weak," "weak," "moderate," "strong," "very strong," and "extremely strong," and the numbers include 0.5 and the integers from 1 to 12. The numbers are spaced logarithmically, whereas the verbal labels are spaced nearly linearly, although with a slight compression toward the top end. The goal of this spacing was to provide a match-up between category labels and numbers that would capture the typical nonlinear relation between category and magnitude-estimation scales (S. S. Stevens & Galanter, 1957).

To each stimulus, whether tone or work level, the subject was instructed to locate the verbal category, or place between categories, that described the experience, and then to report the appropriate number on the scale: Ward (1982) has shown that category rating as well as magnitude estimation can be used in such a multimodal task. (The scale shown in Figure 2 was set out in front of the subject for reference throughout this part of the session.) Interpolation of numbers between stated values was permitted and even encouraged, so the subjects could use decimal values, just as in magnitude estimation. This scale is much like one described previously (Borg, Note 1). It combines the putative ratio properties of magnitude-estimation scales with the ostensible naturalness of scales that employ verbal categories. Moreover, defining the scale in terms of maximal sensation for each subject should serve, according to Borg's (1961, Note 1) model, to make the judgments of exertion comparable across subjects.

As in the first part of the experiment, the levels of work were presented in ascending order, one at a time, for judgment. Heart rates were recorded as before. Interlaced between the six

bouts of exercise and after the last bout were presentations of four of the tones. Thus, in the course of this part of the session, the subject judged each of 6 work levels once, and each of the 12 SPLs twice.

Subjects

Thirty men between 20 and 38 years of age (mean = 26) served as subjects. They were primarily university students who were paid for their participation. By and large, the subjects were in excellent physical condition, many engaging regularly in athletics.

RESULTS AND DISCUSSION

The overall findings are summarized in Figure 3, which plots on a single graph geometric means of the judgments of perceived exertion and of the judgments of loudness, as derived from the two scaling procedures. A significant feature of this graph is the adjustment of the horizontal axes for SPL and for power output in watts. By the method of least squares, straight lines were fitted to log loudness judgments vs. SPL and log exertion judgments vs. log power output for both scaling procedures. Then the two log stimulus scales were adjusted by iterative linear transformations (that is, by adjusting both the horizontal position and the size of the log unit) so as to maximize the coincidence of each set of exertion judgments with the corresponding set of loudness judgments; this was determined by minimizing the deviations of the two sets of judgments from a single line. Although the purpose of this exercise was to simplify the presentation of data in Figure 3, that this could be done at the same time for the results of both scaling methods shows the overall consistency between the two sets of data. Whether the method was magnitude estimation or combined category-ratio scaling, it turned out that a power output of 65 W elicited, on the average, the same numerical judgment as did a sound of about 72 dB SPL, and that a power output of 163 W elicited the same judgment as a sound of about 95 dB SPL.

The two methods did not, however, give identical psychophysical functions. These functions may be characterized by straight lines on these double logarithmic coordinates, corresponding to power functions. The loudness function for data pooled over subjects has an exponent of 0.44 by magnitude estimation and an exponent of 0.36 by category-ratio scaling. Note that the slope of the straight line obtained by magnitude estimation is greater than that obtained by the category-ratio method; the difference between slopes is reliable [$t(29) = 3.45, p < .01$] (see the analysis below and the individual loudness exponents in Table 1). Despite the fact that the category-ratio scale was constructed with the aim of producing a stimulus-response function like that obtained with magnitude estimation, the category-ratio method did give a smaller slope (a power-function exponent 20% lower), as is common with category-scaling methods (Marks, 1968; S. S. Stevens, 1971). On the other hand, note that the average category-ratio

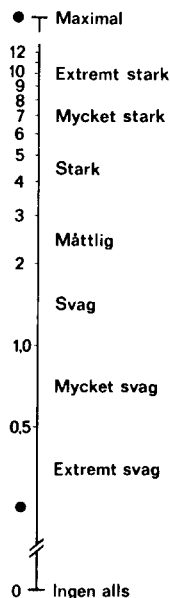


Figure 2. The combined category-ratio scale, visible to the subject during the appropriate part of the test session.

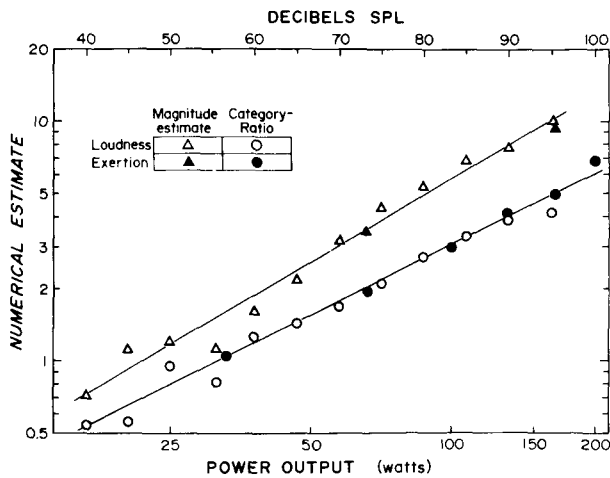


Figure 3. Geometric means of the magnitude estimates (triangles) and category-ratio estimates (circles) of loudness (open symbols), plotted against sound pressure level, and of perceived exertion (filled symbols), plotted against power output in watts.

scale exponent for exertion (mean = 1.53, median = 1.36; see Table 1) is considerably larger than the 1.0 obtained with the RPE scale and reasonably close to the 1.6 typical of magnitude estimation. Furthermore, the internal consistency of the two methods is good: The average matching relationship between exertion and loudness judgments seems to be much the same with the two scaling methods (given that only two points were obtained for perceived exertion using magnitude estimation).

The exponents of power functions fitted to the data provide a convenient summary statistic to evaluate the results for individual subjects. Parameters of individual psychophysical functions appear in Table 1. To each subject's data, we fitted power functions either of the form

$$L = k_1 P^a,$$

for loudness (L) as a function of sound pressure (P), or of the form

Table 1
Parameters of Power Functions for Loudness (L) as a Function of Sound Pressure P ($L = k_1 P^a$) and Perceived Exertion (E) as a Function of Power Output W ($E = c + k_2 W^b$)

Subject	Loudness				Exertion		
	Magnitude Estimation		Category Ratio		Category Ratio		
	a	r*	a	r*	b	c	r*
1	.51	.868	.22	.842	1.01	.00	.994
2	.28	.958	.33	.965	2.06	.45	.992
3	.36	.977	.36	.981	1.40	.60	.986
4	.50	.991	.58	.982	1.35	.00	.998
5	.72	.925	.43	.948	1.58	1.16	.943
6	.58	.969	.53	.969	3.52	.38	.995
7	.36	.996	.42	.968	1.14	.05	.998
8	.36	.926	.33	.966	1.09	.38	.995
9	.40	.990	.33	.987	1.43	.00	.996
10	.47	.992	.35	.974	2.49	1.90	.995
11	.52	.986	.33	.962	1.22	.16	.998
12	.47	.971	.44	.948	1.56	.80	.934
13	.58	.980	.21	.880	1.42	.00	.995
14	.33	.979	.42	.977	1.30	1.20	.996
15	.34	.945	.31	.973	1.61	.70	.991
16	.49	.982	.34	.963	.99	.58	.997
17	.73	.972	.41	.978	1.71	1.12	.998
18	.52	.952	.29	.983	1.12	.00	.983
19	.43	.969	.40	.978	1.13	.00	.995
20	.32	.949	.29	.960	1.35	.84	.993
21	.32	.951	.35	.967	1.14	.42	.983
22	.38	.936	.41	.975	1.60	1.14	.977
23	.52	.988	.25	.881	1.28	1.62	.998
24	.38	.957	.31	.886	1.37	1.24	.993
25	.34	.970	.40	.957	1.18	.22	.995
26	.40	.924	.23	.923	1.94	.00	.983
27	.42	.974	.32	.971	1.52	.46	.992
28	.44	.992	.46	.987	1.20	.02	.999
29	.32	.909	.25	.935	1.06	.00	.992
30	.34	.965	.37	.985	3.01	.94	.987
Mean	.44		.36		1.53		
Standard Deviation	.11		.085		.57		
Median	.41		.35		1.36		

*Pearson r between log L and log P or between log (E-c) and log W.

$$E = c + k_2 W^b,$$

for perceived effort (E) as a function of power output (W). The additive constant c in the latter equation, always estimated as a positive value, reflects at least two factors—a basic level of perceptual noise and the ineluctable perceived exertion involved in pedaling even when there is zero external load. The perceived level corresponding to c typically equals about 4% of maximum (Borg, Note 1). The need for an additive constant shows itself graphically as a slight positive acceleration when the judgments of exertion are plotted in double logarithmic coordinates (see the filled symbols in Figure 3). Obviously, we did not attempt to derive parameters for the magnitude estimations of exertion, given that judgments were obtained at only two levels of stimulation.

Comparison of Magnitude-Estimation and Category-Ratio Scales

As Table 1 shows, the two scaling methods, magnitude estimation and combined category-ratio scaling, yielded essentially uncorrelated estimates of the exponent of the loudness function. Subjects who gave large (or small) exponents by magnitude estimation did not necessarily give large (or small) exponents by the category-ratio method. The correlation coefficient between each subject's two exponents, computed over the 30 subjects, was only 0.178, a small and statistically unreliable value.

This outcome is somewhat surprising, to say the least, given the well-known tendency for an individual subject's exponent to remain rather stable, particularly over a short period of time (see Teghtsoonian & Teghtsoonian, 1971). There is even a clear tendency for the size of the exponent found in one modality to correlate with the size of the exponent in another (Ekman, Hosman, Lindman, Ljungberg, & Åkesson, 1968; Jones & Marcus, 1961; Jones & Woskow, 1962); that is, subjects tend in general to give systematically large or systematically small ranges of numbers. Even this did not happen here to any substantial extent: The correlation coefficient between the loudness exponents and the exertion exponents obtained by the combined category-ratio method was itself just 0.281, also an unreliable association.

The present findings—of virtually no correlation either between the two loudness exponents or between the loudness and exertion exponents—suggest that the two scaling procedures invoked somewhat different kinds of number behavior on the subjects' part. In particular, it appears that the constraints imposed by the verbal categories in the combined category-ratio procedure were sufficient to engender in all subjects a more or less common mode of numerical judgment, and that these constraints were sufficient to overcome the sort of "personal equation" in the use of numbers that each subject brings to the freer magnitude-estimation task.

That a change in number scale can reduce or eliminate such correlations was recently reported by Teghtsoonian and Teghtsoonian (1983). Note, in this respect, that the standard deviation of the loudness exponents is smaller in the combined category-ratio method than in magnitude estimation (Table 1), an outcome that is consistent with the notion of greater constraint in the category-ratio procedure.

Of course, how to interpret this putative reduction in variability is another matter. If one believes that individual differences evidenced in the magnitude-estimation exponents represent real differences in the processing of loudness information, then the somewhat arbitrary elimination of these differences through the combined scaling method would merely serve to obscure important interindividual distinctions. If, on the other hand, one believes that the individual differences among magnitude-estimation exponents largely represent differences in the ways that subjects use numbers—to us, a more plausible interpretation—then the constraints of the combined category-ratio method may serve the useful function of helping to eliminate some of the unwanted variability that results from individual differences in numerical "response bias."

Relation Between Perceived Exertion and Heart Rate

One of the main findings concerns the relationship between heart rate and judgments of perceived exertion. On the average, the heart rate grew linearly with power output, as shown in Figure 4. Heart rate provides a convenient physiological measure of the stress induced by work, a measure that has been shown to correlate closely with ratings of perceived exertion (ratings on the RPE scale; see Borg, Note 1), at the individual as well as at the group level. The correlation between heart rate and RPE goes beyond the fact that both responses increase more or less linearly with power output in cycling. For it is also the case that, at any fixed level of power

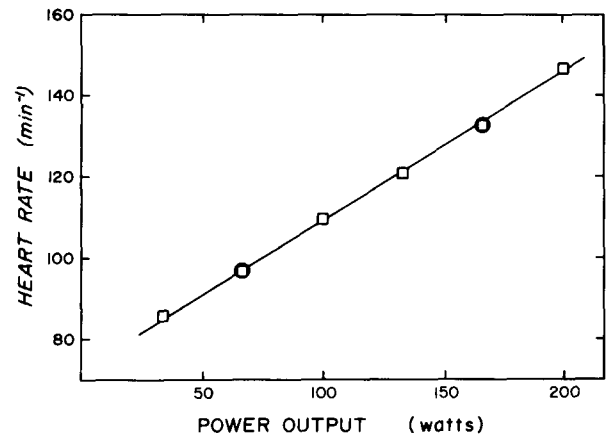


Figure 4. Mean heart rates plotted against power output in watts.

output, the exertion ratings correlate across subjects with heart rate. In other words, those subjects in whom the exercise produces relative high (or low) heart rates also give relatively high (or low) judgments of exertion.

This correspondence appeared in the present experiment: Table 2 gives the correlation coefficient calculated at each power level between the heart rates and the judgments of perceived exertion of the 30 subjects. At the high levels in particular, there are significant and substantial correlations between heart rate and perceived exertion as judged on the category-ratio scale. If a given person's heart rate is relatively high after a bout of cycling, his rating of exertion tends also to be high; if his heart rate is low, his rating of exertion tends to be low. The values of the correlation coefficient are similar to those reported by Borg (Note 1) for judgments made on the RPE scale.

What do these correlations mean? If we assume that an individual's heart rate while bicycling is a measure of the stress on that individual—the greater the heart rate, the greater the stress—then heart rate may also be an indicant of perceived exertion. That one and the same work level can induce in various individuals different degrees of stress and different levels of perceived exertion suggests a theoretical basis for the positive correlation between heart rate and perceived exertion, given that the various subjects make their judgments on the same scale. Magnitude estimation, a method that permits an arbitrary scale factor, does not directly allow comparisons of absolute perceived level across subjects: The fact that two subjects assign different numerals to the same stimulus level need not necessarily imply any difference at all between the sensations experienced; it may merely reflect a difference in the numerical modulus used for judgment. Hence, the low, nonsignificant correlations between heart rate and magnitude estimates of exertion presumably reflect the fact that although there are differences in perceived exertion among individuals in their response to a given stimulus level, these differences are largely obscured by individual variation in choice of numerical scales. Teghtsoonian et al. (1981) considered a similar argument in explaining the lack of invariance across subjects in dynamic range of perceived exertion, as assessed by magnitude estimation. (That there are small positive correlations at all, even if not significant, implies perhaps some slight

tendency for the individual differences in perceptual experience to come through anyhow; this suggests some tendency for different subjects to be able to make absolute judgments of sensation, as Zwislocki & Goodman, 1980, have proposed.)

On the other hand, the correlations between heart rate and category-ratio judgments are good. (Perhaps the correlations would be better were it possible to take into account individual differences in range of heart rate, in interpretation of the category labels, and in "adaptation level" upon entering the experiment. All of these factors could serve to diminish the correspondence between heart rate and judgment.) This outcome is consistent with the following pair of hypotheses: (1) Heart rate correlates with perceived exertion; and (2) perceived exertion is calibrated to maximum sensation at maximum capacity, as Borg (1961, Note 1) has proposed.

Evaluation of Individual Differences by Magnitude Matching

The final main finding was that the method of magnitude matching can serve to differentiate individual functions for perceived exertion, much as the direct assessment by category-ratio judgments can, although perhaps magnitude matching does not do quite as well here as do the direct assessments. First, recall that, on the average, both magnitude-estimation and category-ratio scales of loudness and exertion align themselves well. As Figure 3 shows, the judgments of exertion and the judgments of loudness come together in the same way with both scaling methods. With both methods, the range from 33 to 196 W is, in terms of perceived exertion, equivalent to the range 57-100 dB SPL, in terms of loudness. This agreement holds also at the individual level; if we calculate the SPL equivalent to each subject's judgment of exertion at each work level (see below), we find significant (though not overwhelming) agreement between the results for the two scaling methods: The correlation coefficients (r) between derived values of SPL across subjects are +.380 (65 W) and +.402 (163 W), both values being statistically reliable ($p < .05$). In simpler language, if a given subject "matches" a high (or low) SPL to his perceived exertion by magnitude estimation, he will tend, reliably, to "match" a high (or low) SPL by the category-ratio procedure.

We are also able to evaluate individual psychophysical functions for perceived exertion by magnitude matching. This evaluation, however, is somewhat complicated. To employ magnitude matching, we must transform the numerical judgments into derived cross-modality matching functions. The goal is to compute, for each work level, the SPL that gives a loudness equal to the perceived exertion. This was done following the method described by J. C. Stevens and Marks (1980): The procedure is illustrated by Figure 5, which gives results of one subject. First, straight-line segments connect consecutive points (open circles) on a plot of log loudness versus SPL. Next, we determine the magnitude estimate

Table 2
Correlation Coefficients (r) Between Heart Rate and Judgments of Perceived Exertion, for 30 Subjects, for Magnitude Estimation (ME) and Category-Ratio Scaling (CR)

Method	Power Output (Watts)					
	33	65	98	131	163	196
ME		.099			.236	
CR	.290	.501*	.750*	.664*	.660*	.669*

* $p < .01$.

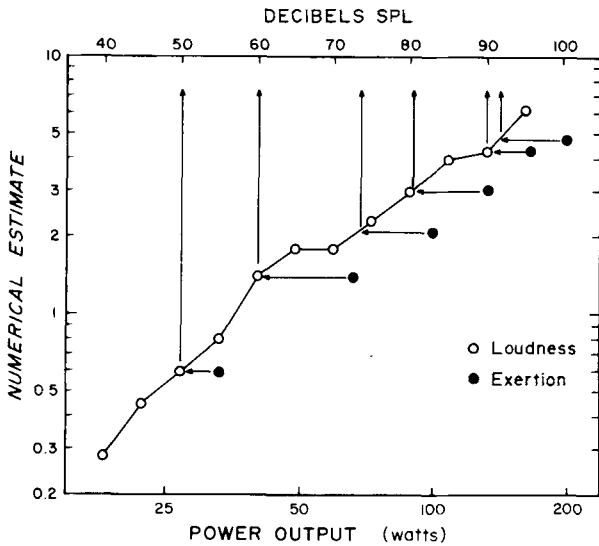


Figure 5. Category-ratio estimates of loudness and of perceived exertion for one subject. The graph shows the method of magnitude matching. For each level of power output, it is possible to find the sound pressure level that gave a loudness equal to the perceived exertion.

or (here) the category-ratio estimate of perceived exertion (filled circles) for each level of power output. Then we project each estimate of exertion horizontally onto the loudness function to locate the loudness that equals the exertion. By projecting upwards, we find the SPL that gave, or would have given, this numerical estimate. In this way, it is possible to derive for each subject a cross-modality matching function that relates SPL to power output (actually, we derive two functions, one for each of the two scaling procedures).

The average cross-modality functions appear in Figure 6. The need for an additive constant in the psychophysical power functions for perceived exertion makes itself evident in the cross-modality matching function, too. Here, the curvature in the function (derived from category-ratio scaling data) was taken into account by estimating the additive sound level (approximately the value of a tone of 50 dB SPL). The exponent of this power function (relating sound pressure to watts) was 3.43 (2.98 for the two points derived from magnitude estimation). These values are a bit high, if one predicts the exponent of the matching function to be 2.67 on the basis of the ratio of the normative exponents for exertion (1.6) and loudness (0.6). If the "true" loudness exponent is taken as 0.6, the cross-modal functions imply an exponent for perceived exertion of 2.0. On the other hand, if the "true" exponent for exertion is taken as 1.6, these cross-modal functions imply loudness exponents of 0.47.

The main concern here, though, is not the slope of the cross-modality function, but the utility of the function in assessing perceived exertion. To what extent do

the individual cross-modality matching functions provide a measure of differences in perceived exertion? Again, heart rate gives a means for comparison. We made the comparisons in two ways. In the first, we utilized the matching values of SPL. In the second, we transformed SPL into loudness, using for this purpose the average loudness functions, as depicted in Figure 3; this was done to bypass the possible effect of the nonlinearity that exists between loudness and SPL. (There is some slight nonlinearity still, because heart rate increased linearly with power output, whereas perceived exertion increased nonlinearly. However, the degree of nonlinearity should be the same when using loudness, computed from the magnitude matches, and when using judgments of exertion themselves.)

Table 3 gives the correlation coefficients, across subjects, between heart rate and both loudness and SPL at each power level. Two features are noteworthy: First, there are significant or near-significant correlations with data derived both from magnitude estimates and from category-ratio judgments, although the latter perhaps gave slightly better values overall. (It is curious, and not easily explained, why SPL yields better correlations than does loudness L.) Second, the present correlations are less good than those between heart rates and direct category-ratio judgments of perceived exertion (compare Tables 2 and 3). This may be because the category-ratio method is intrinsically and in principle superior to magnitude matching at elucidating individual differences in this particular task, namely in judging perceived exertion—or perhaps because magnitude matching necessarily involves two steps, and the use of a second modality in this way unavoidably adds variability to the final assessments. As already mentioned, the

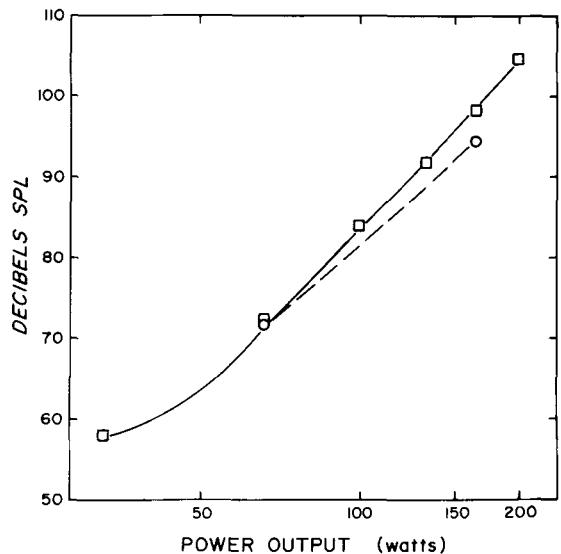


Figure 6. Average cross-modality matching functions between loudness and perceived exertion, determined by magnitude matching.

Table 3
Correlation Coefficients (r) Between Heart Rate and Judgments of Perceived Exertion Transformed by Magnitude Matches Into the Equivalent SPL of a 1000-Hz Tone and Into the Loudness (L) Values of Those 1000-Hz Tones, for Magnitude Estimation (ME) and Category-Ratio Scaling (CR)

Method	Power Output (Watts)					
	33	65	98	131	163	196
	SPL/L					
ME		.382*/.366*			.325 / .232	
CR	.058/.054	.200 / .116	.548**/.468**	.499**/.291	.517**/.339	.491**/.302

* $p < .05$. ** $p < .01$.

derivation through magnitude matching assumes equivalence in the loudness function; yet some interindividual variation in loudness seems likely, and, together with the variability of adding a second perceptual process, perhaps accounts for the smaller correlations.

Given that subjects have normal hearing, that maximal loudness is roughly constant over individuals, and that subjects are able to make "absolute" comparative judgments of loudness and exertion, the positive correlations between heart rate and loudness in our view help buttress the theoretical stance taken here. For instance, as a reviewer of this paper indicated, it is conceivable that the positive correlations between heart rate and categorical judgments of exertion stem in some measure from the intrinsic structure of the task, namely from requiring subjects with different capacities for physical work to judge exertion on the same scale (with fixed numerical constraints). Magnitude matching in our view may help circumvent this possibility. Given the assumptions above, the magnitude matching procedure should eliminate potential artifacts due to constraints imposed by the response scale; the positive correlations between matching loudness and heart rate, then, presumably reflect the tendency for maximal perceived exertion to be roughly constant over individuals.

Be this as it may, it is reassuring to find that both magnitude matching and category-ratio scaling of perceived exertion can yield moderate to high correlations with heart rate, an independently determined measure of the stress involved in physical work. To the extent that this is so, and that heart rate, exertion judgments, and magnitude-matching functions may be taken as differential indications of stress and subjective exertion, the results (1) suggest that both methods can be used to evaluate, statistically, individual differences in perceived exertion and (2) support Borg's notion that, at least as an approximation, the perception of exertion during dynamic work is "calibrated" in each individual according to his work maximum, in that all people experience about the same degree of exertion when they work at their maximum.

REFERENCE NOTE

1. Borg, G. A ratio scaling method for interindividual comparisons. *Reports from the Institute of Applied Psychology*, The University of Stockholm, 1972, No. 27.

REFERENCES

- BORG, G. Interindividual scaling and perception of muscular force. *Kungliga Fysiografiska Sällskapets i Lund Förhandlingar*, 1961, 31, 117-125.
- BORG, G. *Physical performance and perceived exertion*. Lund: Gleerup, 1962.
- BORG, G. Perceived exertion as an indicator of somatic stress. *Scandinavian Journal of Rehabilitation Medicine*, 1970, 2, 92-98.
- BORG, G. A category scale with ratio properties for intermodal and interindividual comparisons. In H.-G. Geissler & P. Petzold (Eds.), *Psychophysical judgment and the process of perception*. Berlin, GDR: VEB Deutscher Verlag der Wissenschaften, 1982.
- EKMAN, G., HOSMAN, J., LINDMAN, R., LJUNGBERG, L., & ÅKESSON, C. A. Interindividual differences in scaling performance. *Perceptual and Motor Skills*, 1968, 26, 815-823.
- HELSON, H. *Adaptation level theory*. New York: Harper & Row, 1964.
- JONES, F. N., & MARCUS, M. J. The subject effect in judgments of subjective magnitude. *Journal of Experimental Psychology*, 1961, 61, 40-44.
- JONES, F. N., & WOSKOW, M. H. On the relationship between estimates of loudness and pitch. *American Journal of Psychology*, 1962, 75, 669-671.
- MARKS, L. E. Stimulus-range, number of categories, and form of the category-scale. *American Journal of Psychology*, 1968, 81, 467-479.
- STEVENS, J. C., & MARKS, L. E. Cross-modality matching functions generated by the method of magnitude estimation. *Perception & Psychophysics*, 1980, 27, 379-389.
- STEVENS, S. S. Issues in psychophysical measurement. *Psychological Review*, 1971, 78, 426-450.
- STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, 1957, 54, 377-411.
- TEGHTSOONIAN, M., & TEGHTSOONIAN, R. How repeatable are Stevens' power law exponents for individual subjects? *Perception & Psychophysics*, 1971, 10, 147-149.
- TEGHTSOONIAN, M., & TEGHTSOONIAN, R. Consistency of individual exponents in cross-modal matching. *Perception & Psychophysics*, 1983, 33, 203-214.
- TEGHTSOONIAN, R., TEGHTSOONIAN, M., & KARLSSON, J.-G. The limits of perceived magnitude: Comparison among individuals and among perceptual continua. *Acta Psychologica*, 1981, 49, 83-94.
- WARD, L. M. Mixed-modality psychophysical scaling: Sequential dependencies and other properties. *Perception & Psychophysics*, 1982, 31, 53-62.
- ZWISLOCKI, J. J., & GOODMAN, D. A. Absolute scaling of sensory magnitudes: A validation. *Perception & Psychophysics*, 1980, 27, 28-38.

(Manuscript received April 6, 1982;
revision accepted for publication April 29, 1983.)