

On the “absoluteness” of category and magnitude scales of pain

WOLFGANG ELLERMEIER and WOLFGANG WESTPHAL
Physiologisches Institut der Universität Würzburg, Würzburg, Germany

and

MARTINA HEIDENFELDER
Psychologisches Institut der Universität Würzburg, Würzburg, Germany

The concept of “absolute scaling” (Zwislocki & Goodman, 1980) implies that direct judgments of sensory magnitude not only reflect the relative positions of the stimuli being judged, but also permit us to assess level differences in sensation. In order to explore this notion for different scaling methods, in the present investigation we compared magnitude estimation with category partitioning, a verbally anchored categorization procedure, in scaling painful pressure stimuli covering different intensity ranges. The results indicate that when the same stimulus range was presented after 1 week, both methods appeared to be highly reliable, with category partitioning faring somewhat better than magnitude estimation. When the stimulus range was unobtrusively changed between sessions, both methods reflected the within-subjects shift in absolute level. When two different sets of subjects judged the slightly different stimulus ranges, both methods resulted in scale values consistent with absolute scaling, though only category partitioning was sensitive enough to differentiate the two stimulus ranges. The results are discussed in the context of different possibilities of anchoring direct scaling methods in order to obtain “absolute” level information.

In psychophysical scaling, an interesting methodological question is whether the direct scaling methods commonly used have certain “absolute” properties. Such properties are particularly desirable in many applications of psychophysical scaling, in which comparisons are made regarding shifts or level differences in sensation (e.g., in the assessment of hearing impairments, or the measurement of the analgesic effects of a drug). It is strongly debated, however, whether direct scaling methods afford comparisons of this kind (cf. Borg, 1982; Marks, 1988; Mellers, 1983a; Ward, 1987; Zwislocki & Goodman, 1980).

Let us first consider magnitude estimation, the most widely accepted direct scaling methodology. Note that in early versions (see S. S. Stevens, 1956), the subject was presented with a standard, and given a number such as 10 (the “modulus”) to assign to it, an implementation that ruled out absolute judgments by definition. Starting in the fifties, however, Stevens tended to favor “free magnitude estimation,” in which the subject is encouraged to match numbers and sensations freely, without the con-

straints of standard or modulus (cf. S. S. Stevens, 1975), and which, at least in principle, would permit “absolute” measurement.

Nevertheless, S. S. Stevens’s approach was concerned primarily with obtaining an unbiased estimate of the exponent of the average psychophysical function, not with displacements of the function along the y-axis. This emphasis led Borg (1962, 1982) to conclude that magnitude estimation is unfit to reveal level differences between individuals or experimental conditions. To remedy this situation, he recommended to verbally anchor subjects’ numerical judgments by using a combined category-ratio scale.

Only a small group of researchers (see Collins & Gescheider, 1989; Hellman & Zwislocki, 1963; Zwislocki & Goodman, 1980) has maintained that magnitude estimation—if implemented with minimal constraints—involves an “absolute” coupling of numbers to sensations (for a review, see Gescheider, 1988). Among the several pieces of evidence produced in support of this claim, one is particularly relevant to the problem of absolute level measurement: In their Experiment 4, Zwislocki and Goodman (1980) presented alternating sequences of a “low” and a “high” range of sound pressure levels (6–54 dB SL vs. 30–78 dB SL) to two different groups of observers and concluded from their results that “the range location of stimulus intensities has little, if any, effect on AME (absolute magnitude estimation) results” (p. 36). There are several reasons, however, to explore this paradigm further. First, it is methodologically unsatisfactory that Zwislocki and Goodman’s conclusion essentially implies accepting the null hypothesis without providing statisti-

This research was supported by grants from the Wilhelm Sander Stiftung (No. 84.013.2) and Deutsche Forschungsgemeinschaft (WE 213/7-3). Portions of the data were presented at the fourth annual meeting of the International Society for Psychophysics, held at Stirling, Scotland, on July 1–5, 1988. We would like to dedicate this article to Otto Heller, at the occasion of his 65th birthday. We are grateful to Martha Teghtsoonian, Lawrence Marks, and an anonymous reviewer for their helpful comments on an earlier version of this paper. Requests for reprints should be addressed to Wolfgang Ellermeier, Institut für Psychologie, Universität Regensburg, Postfach 397, 8400 Regensburg, Germany; e-mail: ellermeier@vax1.uni-regensburg.dbp.de.

cal criteria for its rejection. Second, it is hard to assess the degree of "absoluteness" or relativity without including other methods for comparison. Finally, several recent studies (e.g., Marks et al., 1988; Mellers, 1983a; Ward, 1987) have yielded significant violations of absolute scaling behavior.

On the other hand, as is demonstrated by Borg's (1962, 1982) call for a combined category-ratio scale, there have been occasional claims for absolute level measurement through category scales as well. This may seem odd in the light of the overwhelming evidence demonstrating all kinds of context effects in category scaling (see Helson, 1964; Mellers, 1983a; Parducci & Perrett, 1971). There is, however, a theoretical framework that specifies conditions for "absolute" category scaling and provides a methodology to achieve it.

This theory, known as "Bezugssystemtheorie" (reference frame theory; for a review, see Zoeke & Sarris, 1983) has its foundations in the work of Witte (1966), who was primarily concerned with the development and structure of what he called "mnestically stabilized reference frames," which supposedly form the basis for the categorizations along quantitative dimensions made in everyday life. Grossly simplified, the theory states that the description of perceptual events occurs within a reference frame, which is determined by the day-to-day experiences with a stimulus dimension, and which becomes stabilized in memory. Most naturally, the theory claims, this description occurs in categories of everyday language. As such, it is communicable and absolute; no reference to a standard is needed, reference is always made to the whole "Bezugssystem," not just to the "stimulus context" given in the experimental situation. Thus, while Helson's (1964) approach emphasizes the variability introduced by contextual stimuli, the Bezugssystem approach stresses the stability of judgment derived from what Helson called "residual stimuli."

Heller (1980, 1985) elaborated this theory by specifying the conditions under which absolute judgments can be made, and under which subjects fall back on making comparisons relative to the laboratory context. The latter occurs in "nonoriented" scaling situations, when the categories offered severely mismatch the stimuli provided (e.g., when there are nine categories to describe five weights, all of which are basically "very light," as in a typical adaptation-level paradigm), or when no sufficient anchoring is provided, as is the case with commonly used numerical rating scales. On the basis of these criticisms, Heller developed a scaling procedure ("category partitioning") that directs subjects to categorize stimuli with respect to the Bezugssystem of their experience with the dimension, and discourages a simple ordering of the laboratory set of stimuli. This is achieved through a two-step procedure, in which subjects first have to name the category to which the stimulus "belongs," and then may fine tune using numerical subdivisions.

Given the different claims made, and the conflicting evidence regarding the "absoluteness" of magnitude esti-

mation, another look at the problem seems warranted. A necessity for such an investigation is, of course, a workable definition of "absoluteness": Consistent with the previous discussion, the term "absolute scaling" shall refer here to a one-to-one mapping of the stimulus levels onto the scale, which, operationally, most investigators have taken to imply robustness toward manipulations of stimulus context.¹

A very straightforward way of testing whether such a one-to-one mapping exists consists of shifting the whole stimulus range up or down the stimulus dimension, as in Zwislocki and Goodman's (1980) experiment. Interestingly, this situation is analogous to what is done in pain scaling when an analgesic drug is tested, and it is often used as a test of validity for a given scaling method. In order to be valid, the scale should reflect the reduced pain intensity experienced under the drug (see, e.g., Gracely, McGrath, & Dubner, 1978).

In the present investigation, two largely overlapping series of painful pressure stimuli were constructed. By switching uninformed subjects from a *high-intensity* series to a *low-intensity* series between sessions, both within-subjects and between-subjects comparisons can be made to test for absoluteness in a weak (intraindividual) and in a strong (interindividual) sense. Furthermore, since no comparison has been made between magnitude estimation and category partitioning, in which verbal anchors supposedly reinforce absolute scaling behavior (Borg, 1982; Heller, 1985), such a comparison was conducted for psychophysical judgments of pain intensity.

METHOD

Subjects

Fifty-two undergraduate psychology majors—roughly one third of whom were male, and two thirds female—in their first weeks of study volunteered to participate in the experiment for credit in a laboratory course. All read and signed a consent form, which explained that they could terminate each stimulus presentation at any time or withdraw from the experiment altogether without penalty. Only subjects who were completely naive with respect to the scaling methodologies used, and who were free of acute pain and of analgesic medication, were run in the experiment.

Apparatus and Stimuli

Pressure stimuli were presented by means of a lever mechanism along which a weight could be moved so as to exert more or less pressure on a blunt rod (diameter: 1.8 mm) that rested on the dorsal side of the middle phalanx of the subject's finger. At adequate pressure levels, this stimulus produces a well-reproducible and fairly naturalistic "aching" pain, which builds over time (Göbel & Westphal, 1987). A motor slowly lowered and released the lever for the 20-sec pressure application, which was used throughout the present experiment, unless, of course, the subject indicated tolerance was exceeded, in which case the experimenter pressed a pedal that immediately withdrew the lever.

To investigate the effect of a stimulus shift, two different geometrically spaced series of four pressure levels each were used in the present experiment: The *high* series consisted of pressures of 750, 940, 1,190, and 1,500 kPa; the *low* series comprised pressures of 600, 750, 940, and 1,190 kPa, and thus was displaced downward by one pressure level.

Procedure

Experimental Design

The experimental design is illustrated in Figure 1: 36 subjects were randomly assigned to one of three stimulus configurations (A, B, or C), and to one of two scaling methods (magnitude estimation or category partitioning). In order to equate the number of subjects used in the inter- and intraindividual comparisons (see Figure 1), an additional 16 subjects were randomly assigned to either Condition A or Condition C (first session only). With that exception, each subject participated in two sessions: In Configuration A, the same series of pressure stimuli (the high series) was judged both times; in Configuration B, subjects were switched to the low series in the second session; and in Configuration C, the two series were presented in the reverse order (see Figure 1).

In each session, which lasted for about 1 h (rest breaks included), subjects judged 24 stimuli—all combinations of the four pressure levels with Fingers 2, 3, and 4 of both hands. These stimuli were presented in random order, with the single constraint that no finger recur within a block of six stimulus presentations, in order to avoid sensitization. For the same reason, after each block, rest breaks of 10 min were given in order to allow complete recovery of the "pressure points." Before actual data collection began, two practice stimuli were presented in order to familiarize subjects with the procedure. The second session had exactly the same format and followed the same stimulus sequence, except that the practice stimuli were omitted. In order to present a rationale to subjects that would neither suggest nor rule out changes in sensation magnitude between sessions, the second session was introduced as necessitated by the

fact that "we cannot go through all combinations of fingers and weights in one session."

Scaling Methodologies

Magnitude estimation. A no-standard, modulus-free version of magnitude estimation was used. The instruction (adapted from Marks, 1974, p. 40) encouraged "absolute judgments" and emphasized the ratio properties of the scale.² To make sure subjects understood the magnitude estimation task, they were (following the recommendations of S. S. Stevens, 1975; Zwislocki & Goodman, 1980) first trained in a preliminary task, which required them to judge nine geometrically spaced line lengths presented in random order. The written instructions for the experiment proper ran as follows:

Your task is to tell me how painful each stimulus was by assigning a number to it that reflects its painfulness. The best way to proceed is the following: Assign to the first stimulus that number that best seems to fit it. To the following stimuli, assign numbers in proportion to the perceived pain intensity. Thus, if the pain is three times as strong as that experienced with a previous stimulus, give it a number three times as high. If it is half as painful, assign a number half as high. You may use any positive number: whole number, fraction, or decimal. Don't worry about running out of numbers, for there is always a larger number than the largest one you have used and likewise always a smaller one than the smallest one you have used so far. In case you do not experience pain at all on a given stimulus presentation, you may express that by saying "no pain."

Category partitioning. The categorization procedure used derives from the work of Heller (1985) and was modified for pain scaling by Göbel, Heller, Nowak, and Westphal (1988). The scale used consists of five verbal categories, which can be further subdivided by numbers (see Figure 2).

The following instructions were used:

You are asked to judge the intensity of the painful stimuli using the scale you have in front of you: This scale is subdivided into 5 categories as we commonly use them in everyday life. Please start out by determining the category into which the painful stimulus falls. Then you may "fine tune" your judgment using the numbers within that category. To give you an example: A stimulus was severely, almost very severely painful. The category "severe pain" comprises the numbers 31 to 40. Due to the tendency toward "very severe pain," fine tuning may eventually result in a number close to the upper category boundary, such as a 38 or 39. Always tell me the crude category as well as the number when you make your judgment.

RESULTS

Judgments of the four stimulus intensities presented in a given session were averaged across the six fingers. Arithmetic means rather than geometric means were computed since subjects occasionally gave judgments of zero to indicate "no pain," and also in order to make data reduction and statistical procedures comparable for magnitude estimation and category partitioning. For the 6 (of 52) subjects who had terminated a stimulus presentation because pain approached their tolerance levels, means were computed from the remaining judgments, or—in a few cases—they were extrapolated for the highest level.

The individual psychophysical functions thus obtained were the basis of all further analyses, which focused on (1) the reliability of the judgments, (2) the effect of shifting stimulus intensity within subjects (intraindividual comparison), and (3) the effect of presenting different intensity ranges to different groups of subjects (interindividual comparison).

		STIMULUS CONFIGURATION					
		A		B		C	
		Session 1	Session 2	Session 1	Session 2	Session 1	Session 2
CP	1500	X—X	X—X	X—X	X—X	X—X	
	1190	X X	X X	X X	X X	X X	
	940	X X	X X	X X	X X	X X	
	750	X—X	X—X	X—X	X—X	X—X	
	600	X—X	X—X	X—X	X—X	X—X	
ME	1500	X—X	X—X	X—X	X—X	X—X	
	1190	X X	X X	X X	X X	X X	
	940	X X	X X	X X	X X	X X	
	750	X—X	X—X	X—X	X—X	X—X	
	600	X—X	X—X	X—X	X—X	X—X	

Within-Subjects Comparisons

Between-Subjects Comparison

Figure 1. The experimental design combines the two methods (category partitioning and magnitude estimation, CP and ME) with three stimulus configurations (A, B, and C). The crosses indicate which four of five pressure levels were used in a given session. In Configuration A, subjects were exposed to the same stimulus range in both sessions; in Configuration B, the intensity was lowered by one step in Week 2, and in Configuration C, subjects were switched from the low to the high stimulus range. The two types of comparisons made (intraindividual, interindividual) between the two stimulus ranges are indicated at the bottom of the figure.

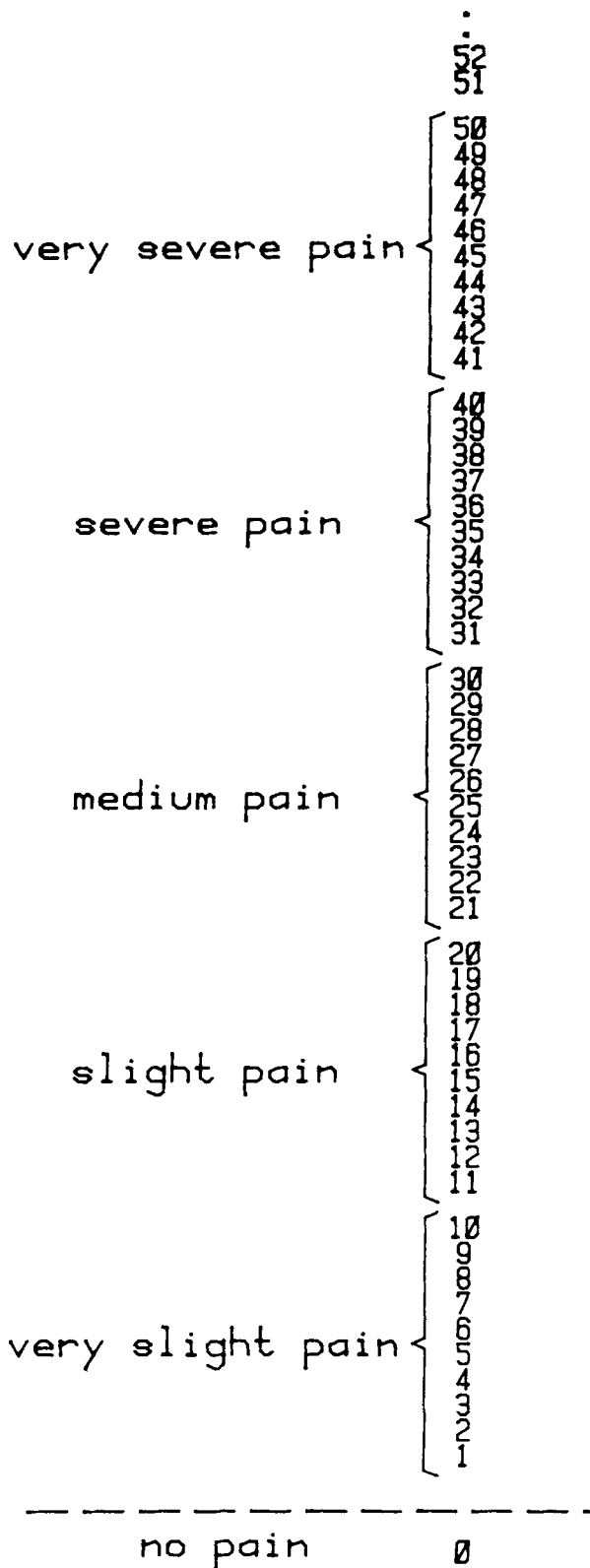


Figure 2. The category partitioning (CP) scale used in the present experiment: Subjects are asked first to determine the verbal category into which the painful stimulus falls, and then to "fine tune," using the numbers.

Reliability

Figure 3 shows the mean pain judgments made by the subjects who received the same four stimulus intensities in their first session and 1 week later. It is evident from the graphs that the mean functions are replicated quite well after the 1-week interval, though somewhat better for category partitioning than for magnitude estimation.

To assess reliability adequately, however, individual subjects' functions have to be inspected. One conventional approach is to correlate a subject's scale values obtained in the first session with those obtained in the second session. With this approach, the average intersession Pearson correlation was found to be $r = 0.9791$ for category partitioning and $r = 0.9671$ for magnitude estimation. Since, however, changes in absolute level (as is evident in Figure 3, bottom graph) would not be reflected in such an analysis, a different approach was taken by computing how much (in percent) the mean judgment made of a given stimulus level changed from Session 1 to Session 2. These absolute changes in scale values computed for each individual subject averaged 9.5% for category partitioning (that is, roughly 3 units on the 50-point scale), and 22.8% for magnitude estimation. A mixed analysis of variance (ANOVA), with the four stimulus levels constituting the within-subjects factor and the two methods constituting the between-subjects factor, indicated that the between-session change in percent was significantly larger for magnitude estimation than for category partitioning [$F(1,18) = 5.21, p < .05$] and was independent of stimulus level.

Intraindividual Intensity Shift

For the intraindividual comparison, the 6 subjects who judged the high series first (Configuration B) were pooled with the 6 subjects who scaled it last (Configuration C), separately for each scaling method, yielding counter-balanced psychophysical functions for the low and high intensity ranges.

Before analyzing the actual data, it might be useful to consider two hypothetical outcomes illustrated in Figure 4. Note that an absolute mapping of stimulus values to responses (Figure 4, top graph) would place all data points on a common function, while maximum relativity would result in two horizontally shifted curves occupying identical response ranges (Figure 4, bottom graph). The actual results are presented in Figure 5, for both magnitude estimation and category partitioning.

The fact that, for both methods, functions of the low and high intensity ranges tend to scatter around a common curve indicates that subjects were able to judge pain intensities in an absolute fashion, with little bias due to stimulus context. Two strategies were chosen to analyze these results:

1. To assess whether the same stimulus intensities elicited the same responses despite different contexts, responses to the three levels common to both stimulus ranges were subjected to repeated measures ANOVAs. The failure to find significant amounts of relativity in these analyses either for category partitioning [$F(1,11) = 1.4,$

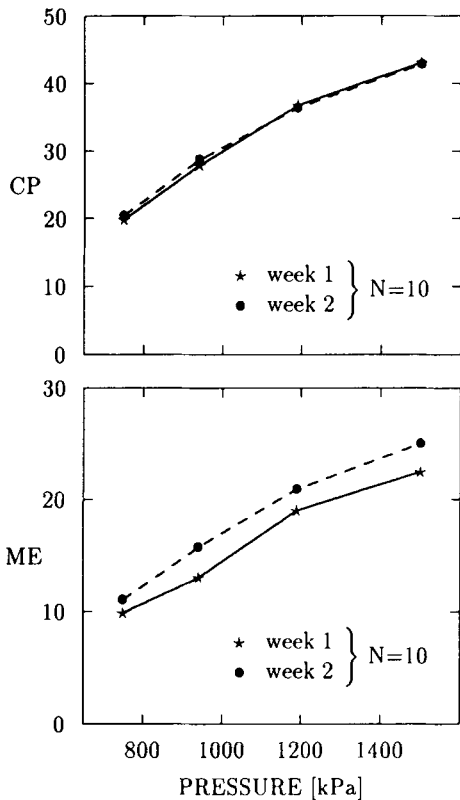


Figure 3. Retest reliability of categorizations (top graph) and magnitude estimates (bottom graph) made by 10 subjects each when the same series of painful pressure stimuli was repeated after a 1-week interval. Stars indicate psychophysical functions obtained in Week 1, circles those obtained in Week 2.

n.s.] or for magnitude estimation [$F(1,11) = 0.12$, n.s.] suggests that subjects largely used the scales in an absolute fashion.

2. A second analysis was conducted to determine whether the two different intensity ranges actually resulted in different pain reports. To achieve this, responses to the four levels of the high series were compared to responses to the four levels of the low series. A repeated measures ANOVA indicated that for both category partitioning [$F(1,11) = 8.35$, $p < .05$] and magnitude estimation [$F(1,11) = 9.22$, $p < .05$], the attenuation (or increase) in stimulus intensity was reflected in subjects' judgments.

While the first strategy may be seen as a test for the absoluteness of the scales, the second strategy assesses their sensitivity to changes in stimulus level.

Overall, this outcome dismisses the argument (most often raised against category scales) that subjects might use the response range available to them like a flexible "rubber scale," putting the strongest stimulus into the highest category, putting the weakest stimulus into the lowest category, and then assigning intermediate positions. The (insignificant) amount of relativity that remains—evident in the rightward shift of the dashed curve in Figure 5—is due to a small number of subjects performing

the task according to a pure "range principle" (Parducci, 1965) by indiscriminately mapping the two different stimulus ranges into the same response range.

The intraindividual comparison, however, tests only a weak criterion for absoluteness—that is, whether judgments of different intensity ranges presented to the same subjects remain consistent over time—and all it requires in the present experiment is some sort of memory for previous judgments, and judging all the stimuli presented in the two sessions on a common scale. Absoluteness in a stronger sense can be tested for, if independent and previously naive groups of subjects judge the two different stimulus ranges. Thus, an interindividual comparison was conducted.

Interindividual Comparison

For this comparison, the psychophysical functions of 10 subjects in each scaling methodology experiencing only the high series (taken from Configuration A, first session) were contrasted with those of 10 subjects experiencing only the low series (taken from Configuration C, first session). These functions are presented in Figure 6.

Again, the data points appear to fall on one common curve. Both methods pass the test for absoluteness, in that mixed ANOVAs of the three overlapping stimulus levels failed to yield significant differences due to stimulus context [$F(1,18) = 1.44$, n.s., for category partitioning;

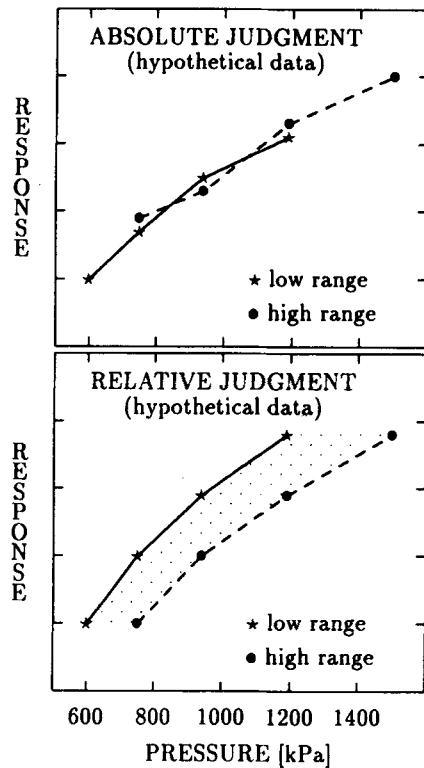


Figure 4. Hypothetical data illustrating an absolute (top graph) and an entirely relative (bottom graph) mapping of stimuli to responses for the present range-shift paradigm. The shading in the bottom graph highlights the displacement of the curves due to relativity.

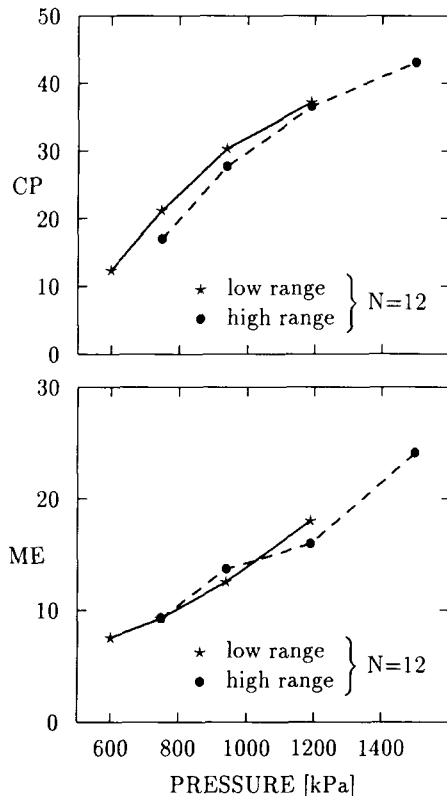


Figure 5. Effect of shifting the stimulus range between sessions for 12 subjects (within-subjects comparison). The top panel shows the results for category partitioning (CP), the bottom panel for magnitude estimation (ME). Circles show mean pain indicated in response to the high stimulus series; stars, the low stimulus series.

$F(1,18) = 0.01$, n.s., for magnitude estimation]. When a test for sensitivity is conducted, however, a different picture emerges for magnitude estimation and category partitioning. As far as the latter is concerned, the scale values clearly differentiate the two intensity ranges, as was confirmed in a mixed ANOVA with the two stimulus ranges and the four stimulus levels as factors [$F(1,18) = 8.9$, $p < .05$]. Free magnitude estimates, on the other hand, did not pass this test for sensitivity. The ANOVA [$F(1,18) = 0.28$, n.s.] showed that responses to the two pressure ranges are not statistically distinguishable (as is evident in the error bars plotted in Figure 6).

DISCUSSION

The results of the present study are quite encouraging as far as the reliability and validity of direct judgments of pain intensity are concerned. Both methods, magnitude estimation and category partitioning, yielded scales that were highly reliable. Both passed the test for "absoluteness," in that the same stimulus levels produced the same responses, whether they were part of a high or a low stimulus range. Furthermore, they proved to be robust toward such manipulations of stimulus context, whether the same subjects served in both conditions or indepen-

dent groups of subjects generated the responses. This outcome suggests that subjects can go beyond a relative mapping of the stimuli immediately at hand into a fixed response range. Of course, such a "no-difference" test for absoluteness will be passed by any scaling method, if only (1) responses vary enough, or (2) the range shift is sufficiently small (imagine shifting stimuli by a quarter decibel in a loudness experiment). The reasoning gains much more power, however, if the test for absoluteness is coupled with a test for sensitivity. In the present experiment, both category partitioning and magnitude estimation turned out to be sensitive enough to pick up the stimulus shift in the intraindividual comparison. When different groups of subjects scaled the two ranges of painful pressure stimuli, only category partitioning was sensitive enough to show a significant change in pain ratings, whereas the magnitude estimation results did not permit a safe conclusion that the two groups of subjects were exposed to different pressure ranges. This insensitivity of magnitude estimation is, of course, largely due to an inherent disadvantage of free or absolute magnitude estimation, the large interindividual variability resulting from the free choice of modulus. Thus, for small- N designs, the category partitioning method used in the present experiment is certainly more efficient.

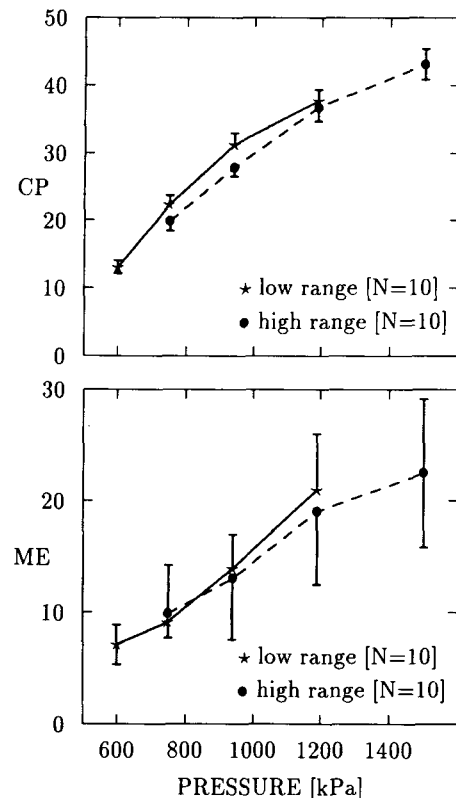


Figure 6. Effect of presenting two different stimulus ranges to different sets of subjects. The graphs show how independent groups of 10 subjects each judged pain produced by the high (circles) and low (stars) pressure series. The top graph shows mean judgments ± 1 SEM for the method of category partitioning (CP); the bottom graph shows the same for magnitude estimation (ME).

The general pattern of outcomes suggests the following conclusions regarding pain measurement. As long as within-subjects designs are used, both magnitude estimation and category partitioning are suitable to measure level effects, such as pain reduction through analgesic medication. The fact that not all subjects, however, perform in accordance with the principles of absolute scaling cautions against making statements about individual subjects (and their responsiveness to the treatment, for example) without having additional information about individual scaling biases. In between-subjects designs, both magnitude estimation and category partitioning would serve equally well in indicating the magnitude of an effect, but with magnitude estimation it might be more difficult to demonstrate its statistical significance.

How do the present results fit into the general picture of studies concerned with the "absoluteness" of magnitude estimation? Whereas some studies provide strong evidence for absolute scaling (e.g., Collins & Gescheider, 1989; Zwislocki & Goodman, 1980), others have found systematic "relativistic" scaling biases (e.g., Marks, Szczesiul, & Ohlott, 1986; Mellers, 1983a; Ward, 1987). On closer inspection, it becomes evident that in all studies done with a range-shift paradigm, both absolute and relativistic components are detectable. This can easily be verified by comparing the amount of relativity actually obtained to the maximum relativity obtainable. Consequently, a 0% shift would indicate truly absolute scaling (as in Figure 4, top graph), a 100% shift a purely relative mapping of stimuli into a fixed response range (as in Figure 4, bottom graph). In Zwislocki and Goodman's (1980) study, the shift is about 15% (i.e., 3–5 dB out of a 24-dB range shift, as read from their Figures 8 and 9); in Marks et al. (1986), it is 59% (see their Figure 1); and even in Ward's (1987) largely critical study, a shift of 33% (i.e., of 3–5 dB out of a possible 12; see his Figure 4) indicates a predominance of absolute scaling behavior. The present study (see Figures 5 and 6), with response shifts ranging from about 5% (for magnitude estimation) to 22% (for category partitioning), is certainly at the better end of these results. It seems that a number of factors, not all of which are understood at present, some being modality-specific (see Marks, 1988; Marks et al., 1986) and some procedural (e.g., the magnitude of the physical shift, time between judgment sessions, and number of repetitions per stimulus; see Ward, 1987), are responsible for the mixed results regarding "absolute scaling."

Beyond contributing to the evidence regarding magnitude estimation, the present study shows that a properly anchored category scale is equally absolute, if not more efficient at measuring level effects. Recent investigations done with range-shift paradigms similar to ours show that conventional numerical (e.g., Ward, 1987) or graphic rating scales (e.g., Marks et al., 1988) fail in this respect. The success of the present categorization procedure may be attributed to two properties that distinguish it from

other category scales: (1) The instructions emphasize that judgments should be made with respect to the "outside-the-lab context" of everyday experience, and reinforce this emphasis by asking subjects always to name the category to which a stimulus "belongs" before further differentiating using numerical ratings. (2) The endpoints of the scale are not only labeled, but operationally defined, in that a pain sensation beyond "50" constitutes the criterion for interrupting a given stimulus presentation, and the zero on the scale corresponds to "no pain" or "just pressure." Borg (1982) has argued that the existence of such well-marked locations as "maximum exertion" serves to provide an absolute anchor that is stable within and comparable across individuals.

Although this may explain why in the present experiment category partitioning did so well, it does not account for the good results of magnitude estimation. It seems that with magnitude estimation there is no well-defined link between points on the scale and sensations. At least two explanations are conceivable: The first one goes back to S. S. Stevens's (e.g., 1975) assertion that magnitude estimation is just another form of cross-modality matching, in this case of numbers to pain sensations, and that the common dimension of intensity serves to link different perceptual continua. More specifically, Zwislocki and Goodman (1980) have argued that it is the concept of numerosity, learned early on in life in the operation of counting, that establishes an "intensity dimension" for numbers leading to their absolute use. On the other hand, it is equally conceivable that the link between numbers and sensations as well as between different perceptual continua is made through a categorical process: "small" numbers are assigned to "small" intensities, "medium" numbers to "medium" intensities, and so forth.

Interestingly, improvements of the magnitude estimation operation have been suggested along both lines (cf. Borg's [1982] combined category-ratio scale; and J. C. Stevens & Marks's [1980] "magnitude matching"). A particularly relevant application of the latter methodology to thermal pain was reported by Duncan, Feine, Bushnell, and Boyer (1988), who demonstrated that magnitude matching distinguished different temperature ranges, whereas traditional magnitude estimation did not. These results, taken together with those of the present study, suggest that the "loose coupling" (Marks et al., 1986) between stimulus and response continua observed in traditional magnitude scaling can be tightened by providing "frames of reference" either through anchored categories (as in Borg's [1982] and in the category partitioning approach) or through cross-modal comparisons (as in the "magnitude matching" methodology). An interesting question for future research might be to investigate whether the judgment processes underlying both strategies are not, in fact, the same. The typical "magnitude matching" instructions (see, e.g., Collins & Gescheider, 1989; Ward, 1987) seem to encourage a matching of categories, and a recent investigation of a number of response scales in taste and

smell (Marks et al., 1988) found that "category matching" was as effective in measuring level-effects as "magnitude matching."

The present investigation demonstrates that category partitioning constitutes one easily implemented option to obtain psychophysical judgments with "absolute qualities." It normalizes different subjects' judgments without wiping out individual differences; it is sensitive even in small-*N* designs; and it contains implicit information about pain and tolerance thresholds.

REFERENCES

- BORG, G. (1962). *Physical performance and perceived exertion*. Lund, Sweden: Gleerup.
- BORG, G. (1982). 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* (pp. 25-34). Berlin: VEB Deutscher Verlag der Wissenschaften.
- COLLINS, A. A., & GESCHIEDER, G. A. (1989). The measurement of loudness in individual children and adults by absolute magnitude estimation and cross-modality matching. *Journal of the Acoustical Society of America*, *85*, 2012-2021.
- DUNCAN, G. H., FEINE, J. S., BUSHNELL, M. C., & BOYER, M. (1988). Use of magnitude matching for measuring group differences in pain perception. In R. Dubner, G. F. Gebhart, & M. R. Bond (Eds.), *Proceedings of the 5th World Congress on Pain* (pp. 383-390). Amsterdam: Elsevier.
- GESCHIEDER, G. A. (1988). Psychophysical scaling. *Annual Review of Psychology*, *39*, 169-200.
- GÖBEL, H., HELLER, O., NOWAK, T., & WESTPHAL, W. (1988). Zur Korrespondenz von Schmerzreiz und Schmerzerleben. *Der Schmerz*, *2*, 205-211.
- GÖBEL, H., & WESTPHAL, W. (1987). Die laterale Asymmetrie der menschlichen Schmerzempfindlichkeit. *Der Schmerz*, *1*, 114-121.
- GRACEY, R. H., McGRATH, P., & DUBNER, R. (1978). Validity and sensitivity of ratio scales of sensory and affective verbal pain descriptors: Manipulation of affect by diazepam. *Pain*, *5*, 19-29.
- HELLMAN, R., & ZWISLOCKI, J. J. (1963). Monaural loudness function at 1000 cps and interaural summation. *Journal of the Acoustical Society of America*, *35*, 856-865.
- HELLER, O. (1980). Orientierung innerhalb phänomenaler Steigerungsreihen. In V. Sarris & W. Lauterbach (Eds.), *Beiträge zur psychologischen Bezugssystemforschung* (pp. 107-136). Bern: Huber.
- HELLER, O. (1985). Hörfeldaudiometrie mit dem Verfahren der Kategorienunterteilung (KU). *Psychologische Beiträge*, *27*, 478-493.
- HELSON, H. (1964). *Adaptation-level theory: An experimental and systematic approach to behavior*. New York: Harper & Row.
- MARKS, L. E. (1974). *Sensory processes: The new psychophysics*. New York: Academic Press.
- MARKS, L. E. (1988). Magnitude estimation and sensory matching. *Perception & Psychophysics*, *43*, 511-525.
- MARKS, L. E., STEVENS, J. C., BARTOSHUK, L. M., GENT, J. F., RIFKIN, B., & STONE, V. K. (1988). Magnitude matching: The measurement of taste and smell. *Chemical Senses*, *13*, 63-87.
- MARKS, L. E., SZCZESIUL, R., & OHLLOTT, P. (1986). On the cross-modal perception of intensity. *Journal of Experimental Psychology: Human Perception & Performance*, *12*, 517-534.
- MELLERS, B. A. (1983a). Evidence against "absolute" scaling. *Perception & Psychophysics*, *33*, 523-526.
- MELLERS, B. A. (1983b). Reply to Zwislocki's views on "absolute" scaling. *Perception & Psychophysics*, *34*, 405-408.
- PARDUCCI, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, *72*, 407-418.
- PARDUCCI, A., & PERRETT, L. F. (1971). Category rating scales: Effects of relative spacing and frequency of stimulus values. *Journal of Experimental Psychology Monographs*, *89*, 427-452.
- STEVENS, J. C., & MARKS, L. E. (1980). Cross-modality matching functions generated by magnitude estimation. *Perception & Psychophysics*, *27*, 379-389.
- STEVENS, S. S. (1956). The direct estimation of sensory magnitudes—loudness. *American Journal of Psychology*, *69*, 1-25.
- STEVENS, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: Wiley.
- WARD, L. M. (1987). Remembrance of sounds past: Memory and psychophysical scaling. *Journal of Experimental Psychology: Human Perception & Performance*, *13*, 216-227.
- WITTE, W. (1966). Das Problem der Bezugssysteme. In W. Metzger (Ed.), *Handbuch der Psychologie I* (pp. 1003-1027). Göttingen: Hogrefe.
- ZOEKE, B., & SARRIS, V. (1983). A comparison of "frame of reference" paradigms in human and animal psychophysics. In H.-G. Geissler, H. F. J. M. Buffart, E. L. J. Leeuwenberg, & V. Sarris (Eds.), *Modern issues in perception* (pp. 283-317). Berlin: VEB Deutscher Verlag der Wissenschaften.
- ZWISLOCKI, J. J. (1983). Absolute and other scales: Question of validity. *Perception & Psychophysics*, *33*, 593-594.
- ZWISLOCKI, J. J., & GOODMAN, D. A. (1980). Absolute scaling of sensory magnitudes: A validation. *Perception & Psychophysics*, *28*, 28-38.

NOTES

1. The term "absolute" is used with quotation marks to distinguish the present usage from a formal definition in terms of scale type (such as ordinal, interval, or ratio). To determine the latter in terms of permissible transformations seems to require more information than can be derived from a single-stimulus task (cf. the debate between Zwislocki, 1983, and Mellers, 1983a, 1983b).
2. This was done to discourage subjects from constructing their own category scale. Note that while Zwislocki and other proponents of "absolute magnitude estimation" have typically avoided reference to number ratios in favor of a pure matching approach (cf. Gescheider, 1988), the experimental evidence suggests that this is not a sufficient condition to obtain "absolute" scale values (see Duncan, Feine, Bushnell, & Boyer, 1988; Marks, Szczesiul, & Ohlott, 1986; Mellers, 1983a; Ward, 1987). The present investigation suggests that it is not a necessary one either.

(Manuscript received April 23, 1990;
revision accepted for publication September 5, 1990.)