Southern Political Science Association

An Application of Psychophysical Scaling Techniques to the Measurement of National Power
Author(s): Allen M. Shinn, Jr.

Source: *The Journal of Politics*, Vol. 31, No. 4 (Nov., 1969), pp. 932-951

Published by: [Cambridge University Press](#) on behalf of the [Southern Political Science Association](#)

Stable URL: <http://www.jstor.org/stable/2128353>

Accessed: 04/02/2014 14:22

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Cambridge University Press and Southern Political Science Association are collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Politics*.

<http://www.jstor.org>

AN APPLICATION OF PSYCHOPHYSICAL SCALING TECHNIQUES TO THE MEASUREMENT OF NATIONAL POWER

ALLEN M. SHINN, JR.

The University of Texas at Austin

INTRODUCTION

In recent years, psychophysicists have expended a great deal of effort in attempts to develop accurate and practical scales of subjective magnitude, i.e., scales that measure perceptions of the magnitudes of "real" variables. These efforts have met with considerable success, at least in the rather limited area in which this work has been done, and it has been more or less conclusively shown that it is possible to measure with considerable accuracy the subjective, or perceived, magnitudes of such variables as loudness of sound, brightness of light, and heaviness of weight.¹

That this is so may not seem of direct relevance to the study of international politics,² but there are indications that the techniques and methodology developed in psychophysics may have direct application to the measurement of social norms, opinions, and attitudes. Thus we have several studies of social status by Hamblin,³ a scale of the seriousness of various acts of juvenile delinquency developed by Sellin and Wolfgang,⁴ a study that scales the importance of Swedish monarchs by Ekman and Kunnapas,⁵ and others.

¹S. S. Stevens, "On The Psychophysical Law," *Psychological Review*, 64 (January 1957), 153-181.

²For an early study that constitutes a notable exception to the generalization implicit in this comment, however, see Frank L. Klingberg, "Studies in Measurement of the Relations Among Sovereign States," *Psychometrika*, 6 (December 1941), 335-352, reprinted in *International Politics and Foreign Policy*, ed. by J. N. Rosenau (New York: The Free Press of Glencoe, Inc., 1961), 483-491.

³Robert L. Hamblin, *Ratio Measurement and Sociological Theory*, Washington University Department of Sociology, St. Louis, n. d. (Mimeographed.)

⁴Thorsten Sellin and Marvin E. Wolfgang, *The Measurement of Delinquency* (New York: John Wiley and Sons, Inc., 1964).

⁵G. Ekman and T. Kunnapas, "A Further Study of Direct and Indirect Scaling Methods," *Scandinavian Journal of Psychology*, 4 (April 1963), 77-80.

The evidence seems to indicate that these techniques may be applicable to the measurement of social values in general, not excluding those that are the concern of students of international politics. It will be the purpose of this paper to investigate, and, if possible, to demonstrate, this applicability in at least one situation. We shall first discuss briefly recent developments in psychophysics that have produced these techniques and then describe an example of their application—an experiment in which we seek to measure perceptions of national power, defined in terms of several “real-world” independent variables. Finally, we shall suggest some additional research areas in which a similar approach may be appropriate.

A major limitation of the study should be noted at the outset. We seek to make a methodological, rather than a substantive, point, and thus we have made simplifying assumptions that may often seem arbitrary and unrealistic. These will become apparent later on as the research is described and discussed, but we ask the reader to bear in mind that it is the methodology, not the substantive findings, that is our real concern.

RECENT PROGRESS IN PSYCHOPHYSICS

Progress in psychophysics has stemmed in large part from a commercial need in the 1930s for a scale that would accurately describe subjective loudness. The need arose because the existing decibel scale, a “real” scale based on sound pressure levels, did not seem to provide a good scale of subjective loudness.⁶ In investigating perceptions of loudness, Stevens and his group presented experimental subjects with a number of different stimulus sound levels, assigned one of the stimuli an arbitrary magnitude to serve as a reference, and asked the subjects to estimate the magnitude of the other stimuli in terms of this reference. The results of repeating this process with perhaps ten to twenty subjects were averaged and then plotted against the decibel values of the stimuli, producing a smooth curve of the form:

⁶S. S. Stevens, “A Metric for the Social Consensus,” *Science*, 151 (February 4, 1966), 530.

$$R = kS^n, \text{ where}$$

R=response value, in arbitrary units

S=stimulus value in decibels

n=an empirical but non-arbitrary exponent

k=an empirical constant dependent on units

The exponent, n , in this case was 0.6, which means that apparent loudness is a decelerating function of sound pressure level (decibels). The plot would look like this:⁷

This technique, known as magnitude estimation, has been applied repeatedly, and experimental evidence is now available to show that for at least two dozen similar continua the same relation in which the response is a simple power function of the stimulus, can be expected to hold. This in turn suggests the validity of a general psychophysical law that equal stimulus ratios produce equal perceptual ratios. This is, of course, a necessary result of the power-function relations.

It appears also that responses to stimuli of this sort are non-voluntary, i.e., they are beyond the conscious will of the subject. This seems almost an obvious conclusion where the continua are those of psychophysics, for which the response appears to be largely physiological. Hence we are not surprised to find only a little variation between individuals with "normal" physical receptors.⁸ But it may also be that responses to social stimuli are non-voluntary in the short run—that an individual is in a sense a pris-

⁷J. C. Stevens, J. D. Mack, and S. S. Stevens, "Growth of Sensation on Seven Continua as Measured by Force of Handgrip," *Journal of Experimental Psychology*, 59 (January 1960), 60-67.

⁸William McGill, "The Shape of the Loudness Function: A Puzzle," in *Psychological Scaling*, ed. by H. Gulliksen and S. Messick (New York: John Wiley and Sons, Inc., 1960), 67-81.

oner of his conditioned values. Given a well-conditioned set of values, his responses are likely to be predetermined; a southern lower-class white responds in a certain way to the stimulus "black skin," and although he may in the long run change his values and thus condition a new response, in the short run his response may depend on the stimulus alone, and not on his free will.⁹ If this is true, then it should be possible to discover power-function relations between stimulus continua describing social values and their perceptions. In fact, this has been done in several studies. Hamblin, in the studies noted previously, related income and education to social status,¹⁰ and in a later study achieved a successful test of the interference - aggression hypothesis,¹¹ finding simple power-function relations in both cases. Sellin and Wolfgang also found such a relation between the seriousness of a crime and the amount of money stolen.¹²

The general technique of magnitude estimation has a number of variations. These are known by such terms as magnitude production or ratio estimation or production, but all are relatively minor variations of the basic approach: the subject is asked to compare two stimuli, one of which serves as a reference, and give an estimate of the relative magnitudes of the two, either in terms of absolute levels of magnitude or in terms of a ratio between them. It makes no difference conceptually whether the subject responds by giving a number, or by squeezing a handgrip, or by drawing a line whose length is proportional to perceived magnitude, or by some other means. All of these techniques may be useful in different experimental contexts, but they all seek to produce the same thing, i.e., a ratio-level scale of the subjective, or perceived, magnitude of some aspect of a group of stimuli. From the means by which they are derived, these scales are known as magnitude scales.

This technique should be applicable to any situation in which the variable of concern can be conceptualized as a perception of a real-world situation, but it may be useful to draw a distinction

⁹Hamblin, *Ratio Measurement*.

¹⁰*Ibid.*, 18-42.

¹¹Robert L. Hamblin, *et al.*, "The Interference Aggression Hypothesis," *Sociometry*, 26 (June 1963), 190-216.

¹²Sellin and Wolfgang, *Delinquency*.

between three classes of such variables, on the basis of variations in the approach to their measurement.

The first of these we may term cultural norms, i.e., values so widely held within a cultural group or subgroup that we can fairly assume that any normal adult member of the group would constitute a competent observer. The phenomenon of social status is such a norm, and in studying its relation to income and occupation, for example, we would seek simply to get a representative sample of the group concerned and would consider each subject, provided he was of sufficient intelligence to understand the questions, to be a competent observer of the phenomenon. The result, then, would simply be the average of the responses of all subjects.

A second class of variables is those for which an observer must have some technical expertise in order to be competent and acceptable for study. The variable "national power" is probably perceived by most people to vary in some way with size and level of development and perhaps with many other variables, but in order to make a worthwhile statement about these relations one must have studied them to some degree. An investigator would be uninterested, probably, in how members of the general public perceive such relations, because the perceptions may well be both inaccurate and little related to other variables, such as policy making. A scholar might be very interested, however, in knowing just how members of the Department of State perceive such relations. In this case the problem is simply to draw one's sample from the group that can be considered to have the necessary expertise. We might call variables of the first two classes "consensus" variables, to emphasize that it is the group response that we seek.¹³

The third class of variables is those that take on unique values for each individual studied, and where it is that value, rather than the group consensus, that is the focus of concern. This class includes, of course, all of what we usually class as attitude or opinion. This class causes some serious problems because, unlike variables of the first two classes, no averaging across individuals is permissible. For the purpose of this paper, however, it is sufficient to note these problems without discussing them, for we shall not deal here with variables of this class.

¹³The term is borrowed from Stevens, "Social Consensus."

Psychophysicists have generally been concerned with variables analogous to those of the first class, for which they hoped to discover stimulus-response relations that could be considered to hold, within certain limits of error, for all "normal" human beings. Their samples have therefore been nonrandom, usually college students, and the results have been averages taken across all subjects in a sample. This averaging procedure seems to be justified, for we would not expect, on the basis of common sense at least, to get different scales of loudness for different cultural groups; certainly we wouldn't expect, and generally don't get, widely varying results between individuals within a cultural group.

The assumption here is that any variation between individuals is simply measurement error, and thus can be safely ignored and, hopefully, averaged out. Although the measurement error in dealing with social-value continua (i.e., the third class above) is likely to be much larger, and the variation between individuals correspondingly greater, the same assumption can reasonably be made when dealing with variables of the first two classes.

In summary, then, we have techniques that have been found useful in psychophysics for relating perceptions to stimuli. On the assumption that perceptions are necessarily important determinants of human action, it seems reasonable to attempt to measure perceptions of stimuli that may have some relevance to international politics. These techniques, known generally as magnitude estimation, are essentially no more than systematic approaches to the time-honored technique of asking questions and arranging the answers. They draw their value primarily from the attention that is paid to asking the question in such a way as to permit valid inferences to be drawn from the answers.

Let us now see whether this scaling technique can be effectively applied to a practical problem in international politics, that of measuring the "power" of nations.

PERCEPTIONS OF NATIONAL POWER

In order to test this measurement approach, we conducted a simple experiment in which the perceptions of national power held by students in two international politics courses were related to stimuli descriptive of the size of nations, of the level of their

economic development, and of the extent to which national resources were devoted to military purposes (militarization).

It should be noted at the outset that we will avoid any discussion about what power "really" is. These discussions have their place, but it is not here, for we necessarily define the term arbitrarily when we design the experiment. Power for us is whatever the common sense meaning of the word becomes when it is connected to the three concepts of size, development, and militarization. Other variables are excluded from consideration; power thus cannot be a relation, nor can it be ascribed to quality of government, or geography, or national morale, or anything else except our three independent variables. This is, of course, opting for a concept of power as a "quantity" rather than a "relation," and puts us somewhat at odds with many recent theorists on the subject.¹⁴ But this position appears inescapable, for the alternative makes the term virtually synonymous with "outcome" and thus renders it useless to *explain* outcomes. As Blalock has noted, had physicists adopted a definition of power in terms of a relation peculiar to a particular situation, the term would have lost all value as a scientific concept, for generalizations would have become impossible.¹⁵

This definition is quite arbitrary and is concededly open to criticism, but it is important to note that the experiment takes this definition for granted, and thus a criticism of the definition is not a cogent criticism of the experiment's internal validity.

The term "experiment," little used in political science, is used intentionally to indicate the nature of the investigation. The essence of experimentation is control; the experimenter ideally holds constant all variables except those under study and observes the effect of his manipulation of the independent on the dependent

¹⁴See, e.g., Robert A. Dahl, *Who Governs?* (New Haven: Yale University Press, 1961); ———, "The Concept of Power," *Behavioral Science*, 2 (July 1957) 201-215; Peter Bachrach and Morton S. Baratz, "Decisions and Non-Decisions: An Analytical Framework," *American Political Science Review*, 57 (September 1963) 632-642; ———, "Two Faces of Power," *American Political Science Review*, 56 (December 1962) 947-952; and Steven J. Brams, "Measuring the Concentration of Power in Political Systems," *American Political Science Review*, 62 (June 1968) 461-475.

¹⁵Hubert M. Blalock, Jr., "The Measurement Problem: A Gap Between the Languages of Theory and Research," ch. 1 in *Methodology in Social Research*, ed. by H. M. Blalock and Ann B. Blalock (New York: McGraw Hill, 1968), 5-27. The relevant discussion is on pages 17-18.

variables. He then seeks to describe the relations, and if his description is accurate, i.e., if his theory is correct, his observed data will fit his predictions exactly. There is therefore an objective criterion for judging his success—the degree of fit between his data and his theory. This degree of fit may be expressed as a correlation, and the goal thus becomes a correlation of 1.0, indicating perfect fit. Anything less than 1.0 indicates a less than perfect experiment, the failure being due, logically, to three possible factors. The first of these is correctness of the model; predictions drawn from a model that is incorrect cannot be expected to fit observed data, although different models may give only slightly differing predictions, and it may therefore be difficult to discover a subtle error here.

The second factor is accuracy of measurement; any inaccuracy in measuring either independent or dependent variables will reduce the degree of fit between predictions and observations. Since we must always expect some measurement error, we can never expect, except by chance, a perfect fit between data and predictions. If measurement is poor it may be impossible to distinguish even between models yielding considerably different predictions.

The third factor has to do with the quality of the experimental design: have all the relevant variables been included in the model or successfully controlled? A variable that is neither controlled nor included in the model will cause unpredictable variation in the dependent variables and will thus decrease the degree of fit between prediction and observation.

The design of this experiment was simple and was intended to satisfy these three points. The variable considered was a “consensus” variable of the sort requiring technical expertise on the part of the subjects. The subjects chosen were students, both advanced undergraduate and graduate, in two courses in international politics at the University of North Carolina. We assume that these students possess the necessary expertise, but we don’t worry greatly over the validity of this assumption, since it affects only the external validity of the experiment, i.e., the ability to generalize its results to other situations, rather than its internal validity. A total of 27 students was used, and acceptable data were received from 25. The remaining two gave data so inconsistent that they were deemed to be worthless and were eliminated from the analysis.

These subjects were given a self-administered questionnaire (copies of which are available from the author), which established a reference point by arbitrarily assigning 100 units of power to a nation of 20 million population (the indicator of size), \$700 Gross National Product per capita (the indicator of development), and which devoted 4% of its GNP to military purposes (the indicator of militarization). The subjects were then asked to estimate the power levels of a number of hypothetical countries, each of which was like the reference country in two of the three independent variables, but different on the third. This procedure amounted to controlling two of the three independent variables while manipulating the third, so that the effect of each of the independent variables could be observed in turn.¹⁶

Note also that this design succeeds in controlling other variables as well, as it must if it is to merit the term "experiment." It does this mainly by definition; "power" is simply that which comes from size, development, and militarization, and nothing else. If we had asked subjects to estimate the power levels of real countries, however, and then correlated the results with their perceptions of size, etc., we would not expect to find that these three variables perfectly determined the result, since clearly a subject could consider other factors as well.

Measurement is the second problem we must solve in order to achieve success, and we did this by averaging across all 25 respondents, on the assumption that differences among them are due to random fluctuations that we call measurement error, rather than to substantive differences that we should deal with explicitly. This assumption ignores the relative expertise of the various subjects; but this will cause us no problems as long as the errors of those with less expertise are essentially random. As we noted, the data provided by two observers were ignored on the grounds that they were so inconsistent as to indicate very low levels of expertise, or no understanding of the problem, or simple unwillingness to cooperate. Even these data would have had little effect on a sample of sufficient size, but with one as small as the sample used in this study it was considered preferable to eliminate them.

¹⁶This design is a slightly modified version of the design Hamblin used in his study of social status, *Ratio Measurement*, 21.

The third problem is to base predictions on a model that is in fact correct, and of course we can never be sure that the model is perfect. One might simply postulate a linear model of the form $P = a + bI$, where P =power, I =the independent variable, and a and b are the regression constants. This is the most common model in the social sciences, because of its simplicity and because the data are infrequently accurate enough to distinguish between this and non-linear models.

The evidence from psychophysics, however, is that perceptions are almost invariably described by simple power functions of the form $R = kS^n$, as noted above. We shall therefore test both models to discover which fits the data better.

As a first step in the analysis each of the independent variables is plotted versus the dependent variable on ordinary ratio graph paper. These plots (Figures 1-3) show a similarity of form, suggesting that power increases with each of the three independent variables, but that the increment of power resulting from each increment of the dependent variable decreases as the values of the independent variables increase. Thus the relations do not appear to be linear (although in the case of GNP/capita versus power the relation is very close to linear), and we can say that power is a decelerated function of each of the independent variables. This is in accord with common sense, of course, as well as with the notion of diminishing marginal returns.

To see whether the data fit the power function model better, we also plot the independent and dependent variables on log-log graph paper, as in Figures 4-6. We should expect to get essentially straight lines if the power function model is correct, because we are essentially plotting the logarithms of the variables on ratio paper, and we know that if

$$\begin{aligned} R &= kS^n, \text{ then} \\ \log R &= \log k + n \log S, \end{aligned}$$

which is an equation of the familiar linear form.

We do find that the resulting curves are very close to straight lines, although apparently not perfectly straight, either because of measurement error or because the hypothesized power-function

model is not quite correct. We can't be sure of the real reason, but there is one possibility we can check out. The log-log plot assumes ratio measurement, and a close inspection of the ratio plots in the vicinity of the origin (Figures 1a, 2a, and 3) demonstrates that the curves do not quite pass through the origins. In the case of population, the correction necessary to make the curve pass through the origin is (+) 0.3 million, and this *threshold correction* should be added to the population values on the log-log curve. This is a very small correction and may well be an artifact of the limits of the technique. We would really prefer to ignore it, because it appears to be in the wrong direction, but we decided to include it for the sake of consistency in the treatment of the data.

In the case of the GNP versus power curve the apparent threshold correction is (-) \$10/year, a correction that serves to straighten the curve somewhat. The same is true for % militarization versus power, where the correction is (-).35%; that is a sizeable correction and one that serves to straighten the curve rather well.

The degree of fit of the data and the regression slopes of the curves presented in Figures 1-6 are summarized in Table 1. It is immediately apparent that all of the correlation coefficients are very high, which is gratifying, but the small differences that exist indicate that the findings are not as clear as we might like. In the case of population, it seems certain that the power-function model is the better predictor, as shown by the difference in r 's: .964 versus .998. The coefficient of determination, r^2 is a more sensitive indicator and shows this difference even more clearly. Similarly, for militarization we would prefer the power-function model, although the difference is not as clear. For development, however, the linear model seems slightly better, and we might be led to choose it were it not for our desire to be consistent and the extent to which the power-function model has been proved by other research. The differences are, in any case, very slight. We prefer, then, to regard national power as a simple power function of development, albeit one that is very close to linear. (As indicated by the exponent of .99. An exponent of 1.0 would be that special case of a power function that is linear.) Note also that Table 1 shows slightly better results when the threshold corrections are applied in two of the three cases, a result we would expect from our discussion of the plots.

TABLE 1
SUMMARY OF SIMPLE POWER RELATIONS

Linear Models		
$P_s = 82.6 + 1.69 \text{ (POP)}$	$r = .964$	$r^2 = .928$
$P_d = 7.04 + .13 \text{ (GNP)}$	$r = .998$	$r^2 = .996$
$P_m = 36.63 + 15.58 \text{ (MIL)}$	$r = .991$	$r^2 = .982$
Power Function Models (No Threshold Corrections)		
$P_s = 8.3 \text{ (POP)}^{.79}$	$r = .998$	$r^2 = .996$
$P_d = .12 \text{ (GNP)}^{1.02}$	$r = .994$	$r^2 = .988$
$P_m = 30.0 \text{ (MIL)}^{.83}$	$r = .994$	$r^2 = .988$
Power Function Models (With Threshold Corrections)		
$P_s = 6.9 \text{ (POP} + 0.3)^{.83}$	$r = .995$	$r^2 = .990$
$P_d = .15 \text{ (GNP} - 10.0)^{.99}$	$r = .995$	$r^2 = .990$
$P_m = 40 \text{ (MIL} - .35)^{.72}$	$r = .998$	$r^2 = .996$

Although we can't at this point explain the slight curvature remaining in the plots, we should not be disappointed by the results so far. If the object of an experiment is to explain the observed variation in the dependent variable, we have succeeded rather well, for in no case does our power-function model with threshold correction explain less than 99% of the observed variation, as indicated by the values of r^2 . This does not by any means guarantee that our model is correct, but it is a far cry from the usual situation in social-science research.¹⁷

We thus have what seem to be acceptable expressions for the relations between perceptions of power and each of the three independent variables, with both of the other variables controlled in each case. We would like, however, to be able to combine these results so as to arrive at a single summary expression for power, and we do this with data gathered in the second part of the questionnaire. In the first part we asked the subjects to estimate power in cases in which only one independent variable was allowed to vary. In the second part we allowed all three to vary and asked for a

¹⁷By comparison, Miller found the average explained variation in a group of studies published in the *American Sociological Review* over a three-month period to be just over 10%. L. K. Miller, *Determination vs. Risk*, Washington University Social Science Institute, St. Louis, n.d. (Mimeographed.) Cited in Hamblin, *Ratio Measurement*.

summary rating based on all three considered together. The results should give us an indication of the relative importance of the three factors, i.e., whether or not one of the factors tended to dominate or overwhelm the others, and, if so, to what extent. Conceptually the dependent variable is total power (P_t), and the independent variables are power due to size (P_s), development (P_d), and militarization (P_m). We might diagram the model as follows:

In the first half of the experiment we concentrated on the left hand side of the diagram; now we shall concentrate on the right side.

We may again hypothesize two possible models (among any number of possibilities) as being plausible. The first would be a simple linear and additive model of the form:

$$P_t = a + b_1 P_s + b_2 P_d + b_3 P_m.$$

We could then use multiple regression to estimate the values of a , b_1 , b_2 , and b_3 . We might also, however, hypothesize a generalization of the power-function model, which would be of the form:

$$P_t = k(P_s)^{n_1}(P_d)^{n_2}(P_m)^{n_3},$$

and would assume a multiplicative relationship in which each independent variable was related to the dependent variable in a non-linear manner.¹⁸ Because of the widespread applicability of the power-function model in most of psychophysics, this seems at least a good hunch. We can again use a logarithmic transformation to make it possible to handle this equation by ordinary least-squares regression analysis. The actual equation tested, then is:

$$\log P_t = \log k + n_1 \log P_s + n_2 \log P_d + n_3 \log P_m,$$

where the estimated values are $\log k$, n_1 , n_2 , and n_3 .

Since we have moved beyond a two-variable situation, we can no longer display our results on graph paper, and so we must de-

¹⁸This is another suggestion for which I am indebted to Robert Hamblin. *Ratio Measurement*, 29-30.

pend entirely on multiple-regression analysis. The results of this analysis are presented in Table 2.

TABLE 2
SUMMARY OF COMPLEX POWER RELATIONSHIPS

Linear, Additive Model	
$P_t = -39.04 + .19P_s + .97P_d + .21P_m$	$R^2 = .787$
Non-Linear, Multiplicative Model	
$P_t = .11P_s^{.50}P_d^{.82}P_m^{.39}$	$R^2 = .984$

We see immediately from comparison of R^2 's, that the multiplicative model is much better than the additive model, so the latter can be rejected as inappropriate. The multiplicative model indicates quite clearly the relative importance of the three factors, with development, as we might have intuitively suspected, the most potent factor.

This equation describes only the right-hand half of the diagram shown in Figure 7, and we would prefer an expression that describes the entire process by which the factors of size, development, and militarization apparently combine to produce perceptions of power. We can do this by substituting for P_s , P_d , and P_m the expressions given in Table 1 for these quantities (with threshold corrections applied), and if we do this we have the following expression:

$$P_t = .37 (\text{POP})^{.41} (\text{GNP})^{.62} (\text{MIL})^{.28},$$

where each of the independent variables is corrected for threshold. The accuracy of this expression can be estimated by taking the product of the r^2 's of the expressions for the independent variables, and the overall R^2 , which yields a figure of .961, or 96.1 percent accuracy.

It would appear, then, that this technique is appropriate, at least for the continuum under study. By taking data from a group of twenty-five subjects and averaging the results, we were able to derive a single expression that expresses how these subjects perceive power as a function of national size, development, and mili-

tarization, and that does so sufficiently well to account for just over 96% of the observed variation in the data. We should emphasize that this is not 96% of observed variation in the original data, but of the variation in the averages taken across all subjects. For the continuum under study, however, this is not a limitation, since we have assumed already that variation between individuals is due to random-measurement error.

DISCUSSION

If we assume that student perceptions are congruent with those of national elites we can draw a number of inferences from the results of this experiment.

First, we have an excellent idea of the relative importance of our three independent variables in affecting individuals' perceptions of national power. Inspection of the exponents in the final equation for power indicates quite clearly, for instance, that development, in terms of GNP/capita, is far more important than militarization in calculating power. The implication of this finding, if one sought to follow a national policy of increasing power, should be apparent.

Secondly, the shapes of the curves tell us something about the likely effect of an added increment of each of the independent variables. All variables have curves that are concave downwards, so that each added increment has less and less effect on the total power level as the level of the independent variable increases. There is, however, a considerable difference in the degree to which the curves are concave. The development curve is very close to linear, whereas the militarization curve is much steeper near the origin than at higher values, indicating fairly sharply diminished marginal returns at high levels of militarization.

The most important use of these findings, however, will be in further research, when they can be used as a measure of national power in studies relating power to other variables of theoretical interest. This, of course, assumes the substantive validity of the specific findings we obtained, and we believe that this assumption should not be made lightly. First, there are unanswered questions about the value of asking students to make judgments in this area; it might be advisable to seek far more expert judges, closer,

perhaps, to the policy makers themselves. The narrow manner in which we have defined power also presents a problem. Clearly the definition should be broadened, but the other factors that might affect power judgments are unknown, and these can be discovered only through further research. Finally, we would like to see this experiment replicated, with variations in the subjects used, the variables used, and the details of design, so that we might be better able to evaluate the stability of the findings. Conclusive results could easily require a dozen or more replications by different researchers.

Despite all these caveats, some researchers may find the present results to be of value. Particularly in studies of an exploratory nature—and this seems to include the great bulk of current theoretical research in international politics—the expression we have derived for perceived power may be accurate enough to use as is. Where it can be combined with other variables measured on an interval or ratio level, of course, the more powerful statistics associated with this level of measurement can be appropriately used, and this may be a consideration of considerable importance in designing a research project.

CONCLUSIONS

We sought in this project to answer, if we could, the question whether or not the measurement techniques developed in psychophysics have any relevance to the study of international politics, or to social science in general. The experiment just described seems to indicate that there is relevance. Within limits, it seems clear that these techniques will find a place in social-science research.

The most important of these limits is that we must be willing to deal with perceptions. The researcher who deals with aggregate data will find no hope here, and the same limitation applies when the variables, even on an individual basis, are concrete, “real-world” variables such as social background or roll-call votes. But if one is willing to assume that a policy-maker’s perceptions of those real world variables will determine his actions, then perhaps one will find these techniques of some use. And this may not be a difficult assumption to accept. As John Stoessinger writes, “. . . to some degree, at least, international relations are what people think

they are; or, to put it in other words, . . . men respond not to realities but to fictions that they have themselves created.”¹⁹ Ivo Duchacek emphasizes a similar theme when he states, “. . . the image of the world situation that top policy makers have constructed and accepted as objective in their own minds is more important than any other image, including the correct one.”²⁰ In the study of national power we looked briefly at the way in which some students perceived the effect of certain real-world variables. These students’ perceptions, of course, are not particularly relevant to policy making, but if we accept the contention that those perceptions are conditioned and change only slowly with time, it might be very useful to study such perceptions in a group from which top policy makers may be drawn. Students at the National War College, for example, are such a group, and it is not beyond possibility that approval for a study of such a group could be obtained.

Perceptions of the policy makers actually in office, of course, are difficult to study, although the problem is not impossible. In a recent paper, Winham called attention to the possible uses of content analysis in dealing with these problems.²¹ The assumption is that policy makers will reveal their perceptions of a situation in their published communications, a corollary being that the scholar will be able to judge these perceptions accurately by reading these communications. The whole process could be sharpened considerably by the use of psychophysical scaling techniques. If one wished to investigate the relation between perceptions of threat from the communist world and votes in the Senate Foreign Relations Committee, one might select as one’s research technique the analysis of statements of Committee members carried in the Congressional Record. One would then proceed to code these communications according to one’s judgment of the perceived threat level they implied, perhaps as “high,” “medium,” and “low.” This procedure would yield rough measurement at best, but it would

¹⁹John G. Stoessinger, *The Might of Nations* (rev. ed.; New York: Random House, 1965), 395.

²⁰Ivo D. Duchacek, *Nations and Men* (New York: Holt, Rinehart and Winston, 1966), 12.

²¹Gilbert R. Winham, “The Use of Quantitative Indicators in Foreign Policy Analysis” (paper presented to the Canadian Political Science Association, Hamilton, Ontario, McMaster University Department of Political Science, June 1968).

be entirely possible to refine these categories by selecting a group of "expert" judges, who would then rate a number of statements on the continuum, "perceptions of threat." The result would be the assignment of numbers on a ratio scale to each statement, a far more accurate process than the simple assignment of "high," "medium," and "low" categories.

At least two other worthwhile insights for political scientists stem from this general methodological approach. The first of these is that pure experimental research designs *are* possible in international politics and may prove quite valuable in theory construction. This is true primarily because only these research designs permit the degree of control that, when coupled with adequate measurement, will allow the accurate testing of different models. The difference between the two alternative models of national power, for example, which is made obvious by the R^2 's, presented in Table 2, could probably not have been detected, and certainly not so convincingly, if an experimental design had not been used to control other variables.

The second insight stems from the first and leads one to conclude that it may be worthwhile to investigate models other than the usual linear and additive ones. It is not reasonable to assume that much of social reality can be accurately described by simple linear models.

The problems of applying these techniques to social science are by no means solved, and we do not want to minimize the remaining difficulties. Nor will we argue that they are of use in all areas of social science; clearly they are not. Nonetheless we maintain that the basic relevance of these techniques to some of our problems is reasonably well established, and that the problem now is to work out the details of their application. In this manner we hope to advance the science of politics.