



## VARIETIES OF MILITARY LEADERSHIP

Hanan C. Selvin      *State University of New York, Stony Brook*

WORKERS ON an assembly line, students in a third-grade classroom, and soldiers in an army training camp do different kinds of work in radically different settings, but they have in common one important social relationship. They all spend a good part of their day in close contact with lower-level leaders, such as foremen, teachers, and company-level officers, both commissioned and noncommissioned. From both individual experience and empirical research we know that the behavior of workers, students, soldiers, and others in subordinate positions, at work and afterward, is significantly affected by the actions of their leaders.

The empirical study reported here shows how the actions of company leaders in twelve U.S. Army training companies affected the *nonduty* behavior of several hundred soldiers undergoing basic training. The unraveling of these effects of leadership was unusually complex. Unlike the student and the worker, who usually are subject to only one leader in the course of a

working day, the trainee had two company-level commissioned officers (the Commanding Officer and the Executive Officer) and two company-level "non-coms" (the First Sergeant and the Field First Sergeant). There was constant turnover in these positions: during his training cycle the typical trainee had seven company-level leaders.

An additional source of complexity in this study is the way in which we constructed descriptions of the "leadership climates" of the companies. For this study, it seemed better to rely on the trainees' description of their leaders in a questionnaire that they filled out at the end of their basic training, rather than on judgments by superiors or outside experts, as is often done in evaluating how well an organization achieves its goals. Accordingly, each trainee rated each of his leaders on fifteen different questions, ranging from how well the leader inspired confidence to whether he punished the men at every opportunity.

The sheer bulk of these data is impressive: an average of 150 men in each of 12 companies rated an average of 7 leaders on 15 questions. Multiplying these figures together ( $150 \times 12 \times 7 \times 15$ ) yields a total of about 189,000 separate ratings of company leaders. The major statistical problem was to boil down this mass of data into descriptions of the leadership climates of the companies.

Part of this statistical "boiling down," or *data reduction* as it is usually called, consisted of such simple procedures as computing averages. Another large part, much more complex and more illuminating, was a statistical procedure called *factor analysis*, which played a central part in measuring the leadership climates. These factor-analytic procedures not only exemplify a powerful technique, but also can be applied whenever several people can give independent judgments about someone with whose behavior they are familiar. Examples are teachers as described by their students, students as described by several teachers, and mental hospital patients as seen by the ward staff. Finally, although most studies of teaching effectiveness in colleges and universities rely on ratings of the teachers by their students, they do not typically go on to the kind of analytic clarification that this procedure would afford.

#### THE IDEA OF FACTOR ANALYSIS

The three primary colors (red, yellow, and blue) when suitably combined, yield thousands of different colors. Similarly every bit of matter can be analyzed into some combination of the hundred-odd chemical elements. These two facts, familiar to all adults from their school days, are parallels in the realm of physical science to what the statistical procedure of factor analysis can sometimes do with such social phenomena as opinions, votes, and symptoms of mental illness. Factor analysis is, in short, a way to discover or construct

from a larger group of observed characteristics, or *items*, a small set of more general characteristics, or *factors*, various combinations of which will produce each of the observed patterns of items.

#### THE BACKGROUND FOR THE DATA

To explain this work, let us start with the gathering of the data at Fort Dix, New Jersey, in the spring of 1952 by two physicians, Arthur M. Arkin and Thomas M. Gellert, then on the staff of the Mental Health Consultation Service (a central psychiatric facility to which soldiers were referred from dispensaries located near their companies). Over a period of several months they began to notice patterns in their records. Some companies had higher rates of accidents, other companies suffered more psychosomatic illnesses, and still other companies had greater proportions of men going AWOL for short periods. Because all companies followed essentially the same program of training, lived in identical barracks, and ate the same food, the staff members speculated about the kinds of factors that might be responsible for the differences they had observed. They reasoned that differences in the nature of the leadership among companies might account for the differences in rates of accidents, of psychosomatic illnesses, and of going AWOL.

Further reflections and some pilot studies soon led to the development of the two questionnaires that are the basis of this study. One, the "behavior questionnaire," asked each trainee to report the frequency of 24 kinds of nonduty behavior, such as going to the PX for food between meals, having sexual intercourse, going to the movies, and seeing the Chaplain. This questionnaire also asked for the trainee's age, education, and marital status.

The second questionnaire dealt with the company-level leaders that the trainee had had during the sixteen weeks of basic training: Commanding Officer (C.O.), Executive Officer (Exec.), First Sergeant (1/Sgt.) and Field First Sergeant (F-1/Sgt.). In general, the C.O. and the F-1/Sgt. worked directly with the trainees, and the other two leaders usually remained in the company office, or orderly room, and had less contact with the men.

#### APPROACHES TO THE ANALYSIS

Three elements combined to shape the analysis of the leadership data: the nature of the data as described above, the properties of the available statistical methods, and my training as a sociologist. At the outset of the analysis, there was a choice between two essentially different problems: the *psychological* problem of trying to explain a particular event (say, why Pvt. John Doe got drunk on his first weekend pass) and the *sociological* problem of variations in *rates* of behavior in different social units (why, for example, did Doe's

company have a higher proportion of men getting drunk than did any other company?).

I chose to work on the second problem, both because of my training as a sociologist and because the data lacked the detailed psychological information on each individual soldier needed to learn why he behaved as he did. Once made, this decision helped to shape the answer to the second basic question of the study: how to describe the leadership of each company. It gradually became clear to my assistant (E. David Nasatir) and me that the data had to be put together in two different ways. First, we wanted to describe the leadership behavior of all of the leaders in a company, not simply that of the C.O.; we expected the nonduty behavior of the trainees to be affected by the overall leadership climate of the company. (We were able to show that each leader contributed something of his own to that climate and that his actions were not simply copies of the actions of the C.O.) Second, all of the leadership data we had were embodied in the responses of the individual trainees, so it was necessary to combine the responses in some way for two reasons. We wanted to find the common elements in the evaluations of leadership in a company, not the idiosyncratic perceptions of one or a few trainees, and, consistent with our sociological orientation, we wanted to focus on how the trainees in each company, as a group, saw their leaders.

The central statistical task is thus to describe the entire set of company-level leaders as seen by the entire set of trainees in each company, in other words, to reduce the 189,000 ratings of the 82 leaders by the 1800 trainees on the 15 questions to a small set of descriptions of each company's leadership climate.

#### THE IDEAL STUDY AND THE REAL STUDY

It will clarify the statistical reasoning to put the questionnaires aside for a moment and ask how one would go about describing leadership climates if one had unlimited resources of money, trained personnel, and time.

Ideally, perhaps, one would assemble a group of trained observers—or even one “omniscient observer”—and ask them to live with each company for a significant part of its training cycle. These observers would watch, record, and evaluate the behavior of the leaders and somehow produce a concise description of each company's leadership climate.

Even if everything were ideal, this would be extraordinarily difficult. For one thing, the observers would have to be everywhere, watching everything, and yet not interfering with the training activities or affecting the nature of the leadership. No, a corps of observers would not do, but if we could depart altogether from a realistic observational situation, at least to the extent of thinking of what an ideal arrangement might be, we would like to have an omniscient observer, a kind of observational superman who could see every interaction, describe it, and combine it appropriately with all of the thousands

TABLE 1. How Pvt. Doe Answered Question 15 of Leadership Questionnaire

|  | Exec. C.O. | 1st Off. | Field Sgt. | 1/Sgt. |
|--|------------|----------|------------|--------|
| 15. If you were ordered into combat and you could choose the men who would be your leaders use the No. 1 for those men in your unit you would like MOST to lead you; No. 2 for those men whom you would like LESS to lead you; and the No. 3 for those men you would like LEAST to lead you if at all. |            |          |            |        |

of others that he observes. Such an omniscient observer does not exist, but we were able to create an approximation to his observations statistically by basing the descriptions of the leaders' behavior on the experience of the trainees. To see how this was done, consider question 15 (see Table 1) of the leadership questionnaire filled out by Pvt. John Doe of company X.

During his 16 weeks of training, Doe had seven company-level leaders: two C.O.'s, one Executive Officer, two First Sergeants, and two Field First Sergeants. The numbers at the right in Table 1 are his ratings of each officer and noncom as a combat leader. Doe apparently thought that the first C.O. would have made a good combat leader, for he gave him the highest rating, 1. His unwillingness to follow the second C.O. into combat is indicated by the low rating of 3.

Every trainee in Doe's company rated the same company-level officers on this question. For the sake of illustration, assume that there were 100 trainees in this company and that their ratings of the first C.O. as a combat leader were those shown in Table 2. The average of these ratings is 1.70, so this is the rating that the first C.O. was assigned on combat leadership.

TABLE 2. Ratings Given to First C.O. by 100 Trainees

| RATING<br>(1)                      | NUMBER OF TRAINEES<br>(2) | (1) × (2) |
|------------------------------------|---------------------------|-----------|
| 1                                  | 50                        | 50        |
| 2                                  | 30                        | 60        |
| 3                                  | 20                        | 60        |
|                                    | 100                       | 170       |
| Average = $\frac{170}{100} = 1.70$ |                           |           |

We can now turn away from the trainees and take each average rating as a characteristic of the leader being rated. Thus the first C.O. in the illustrative example would be said to have a rating of 1.70 as a combat leader. In other words, the average ratings received by a leader may be considered as his (*perceived*) attributes.

In the leadership questionnaire each leader was rated, as in the foregoing illustration, on the extent that he:

- (1) Influenced the lives of the trainees
- (2) Commanded the respect of the trainees
- (3) Was a "sucker for sob stories"
- (4) Was a "good Joe" one minute and "mean as Hell" the next
- (5) Could create a real fighting spirit against the enemy
- (6) Acted in such a way that the trainees were afraid of him
- (7) Could not be depended on to keep his promises
- (8) Created a feeling of confidence in the trainees
- (9) Told the trainees when he thought that an order from higher headquarters was unfair or silly
- (10) Displayed a real interest in the trainees without babying them
- (11) Treated the trainees "like dirt"
- (12) Gave more breaks to his favorite trainees than to others
- (13) Seized every opportunity to punish his men
- (14) Tried to have his men excused from "dirty details" ordered by higher authorities
- (15) Would be preferred as a leader in combat

This use of average ratings, instead of the original ratings by each of the trainees, yields an impressive reduction in the amount of data. Instead of some 189,000 individual ratings, there are now only 1230 averages (82 leaders rated on 15 questions). Even more important than the quantity, however, is the quality of these statistically derived data. The original ratings of each leader show a great deal of variation, with misperception, failure to follow instructions, facetiousness, and errors of processing all distorting the true ratings.

#### COVARIATION OF RATINGS

The quality of these average ratings appears most clearly when we see how much the characteristics of the leaders that *should* vary together *do* vary together. To show this, we must introduce a numerical measure of this joint variation. We chose the *coefficient of correlation*, invented by Sir Francis Galton in the 1880s to measure how much various physical characteristics,

such as height, are inherited. If the height of a son can be predicted *exactly* by a mathematical equation for a straight line using the height of his father and if tall fathers give rise to tall sons, then the value of the correlation coefficient is 1.0, the largest value that this coefficient can have. If the height of a son can be predicted exactly from the height of his father, but tall fathers produce short sons, then the value of the correlation coefficient is  $-1.0$ , its largest negative value. And, if there were no relation between the heights of fathers and the heights of sons, the correlation coefficient would be 0. (See the essay by Whitney for a further description of the correlation coefficient.)

In real data on individuals, values close to 1.0 or  $-1.0$  are rare. Thus the correlations between pairs of ratings given to any one leader by the men in his company seldom were higher than 0.30. These are, of course, the correlations between the responses of the individual trainees to the leadership questions, before the computation of averages. For example, a trainee who rated a particular leader high as a combat leader might be almost as likely to rate him low in displaying an interest in the trainees as he would be to rate him high on this second trait.

The situation is altogether different for the average ratings. For example, a leader who has a high average rating on instilling a fighting spirit in his men (question 5) almost always has a high average rating on commanding their respect (question 2); the correlation between the averages on these two characteristics is 0.82. Similarly, a leader who punishes at every opportunity (question 13) usually produces fear (question 6); the correlation in this case is 0.84. The size of these correlations between averages is striking: of the 105 correlations in the leadership data, 49 are numerically greater than 0.50, 28 are numerically greater than 0.70, and 13 are numerically greater than 0.80.

On the other hand, we might expect competence and coercion to be negatively related—that, by and large, leaders who were judged to be competent would be less likely to be judged coercive. The data only partially bear out this expectation. The correlations between the average scores on inspiring respect and the averages on the two questions that measure coerciveness (6 and 13) are moderately negative ( $-0.28$  for the question on fear and  $-0.45$  for the one on punishment), but the corresponding correlations between the averages on the question on instilling a fighting spirit and the averages on the measures of coerciveness are so close to zero ( $-0.01$  and  $-0.16$ ) that they indicate that there is no appreciable relation.

Even though there are only four items in the analysis in the previous paragraph, the discussion was a bit complicated. Part of this complexity might be removed by a better choice of words, but there is a limit to the complexity that words can clarify. Imagine the complexity in trying to relate all 105 correlation coefficients between pairs of averages, instead of only four!

## USING FACTOR ANALYSIS TO DESCRIBE THE STRUCTURE OF LEADERSHIP

Factor analysis offers a way out of this complexity. This statistical procedure often makes it possible to untangle large sets of correlation coefficients; in brief it determines which items go together and which do not. Moreover, it expresses this structure of relations numerically, so that we can tell *how much* of what kinds of order there is in the set of correlations and how these simpler orders fit together.

Before turning to the leadership data it is important to say a few words about the goals of factor analysis. In psychology, the field where factor analysis was invented and has been most used, it is customary to speak of the statistically derived factors as "underlying," "basic," or "fundamental" variables and to use the verbs "discover" or "uncover" to describe the process used to calculate the factors. This language suggests that psychologists and statisticians have invented a statistical procedure for discovering scientific truths, much as chemists discovered the 100-odd fundamental chemical elements. I prefer to use a different set of terms. I shall speak of "constructing new variables," or factors, from combinations of the original items. The statistical procedures are the same; only the shades of meaning attached to them differ.

It turns out that the 15 original questions in the leadership data can be combined into 3 new variables or factors, which we labeled "positive leadership," "tyrannical leadership," and "vacillating leadership." For example, a leader who received averages close to 1 on the question of willingness to follow into combat and on similar questions that make up the factor of positive leadership would get a high score on that factor. The factor-analytic computations thus lead to a set of scores for each leader on the 3 factors, scores that, to a considerable extent, can replace his scores on the 15 original variables. That is, if we know a leader's scores on these factors, then we can estimate his average ratings for the 15 original questions with a high degree of accuracy. The statistical relations between items and factors suggest the names for the factors. Thus a leader who scores high on the factor of positive leadership is one who creates confidence in the trainees, is able to instill a fighting spirit in his men, is interested in them, and is one whom the men would like to have leading them in combat. Similarly, leaders who receive high scores on the second factor, tyrannical leadership, are likely to be seen as producing fear in the trainees, as punishing them at every opportunity, and as treating them "like dirt." High scores on vacillating leadership go to leaders who play favorites, punish at every opportunity, and are "Good Joes one minute and mean as Hell the next."

The construction of the 3 factors from the original 15 items was entirely a statistical operation, based only on the numerical correlations among the 15 items. Neither the wording of the questions nor the analyst's expecta-

tions entered into these computations. These nonstatistical considerations enter only in choosing the names for the factors, and even these choices are relatively unimportant when one has access to all the significant numerical results.

## VERIFYING THE MEANING OF THE FACTORS

Instead of dreary columns of numbers, let us look at other, perhaps more meaningful evidence that these factors really do express the trainees' perceptions of their leaders. The evidence comes from the unsolicited comments that many trainees wrote on the leadership questionnaires. For example, one trainee wrote of an officer who turned out to have a particularly high positive-leadership score:

I think that our commanding officer, Capt. ———, was a great leader, he held the respect of all the men and was just about everyone's choice to lead them in combat if we ever saw action.

And a First Sergeant who happened to receive a conspicuously low score on this dimension elicited the remark:

... he is the most unsympathetic character that I have ever encountered in my life also sneaky . . . . I don't see how he ever earned his stripes for he has the mental capabilities of a mongoloid.

Similarly, an officer with a very high score on tyrannical leadership drew this comment:

The C.O. beat men until they ran to the I.G. (Inspector General). Very few of us got passes during basic. We never got breaks in our marches because the C.O. was either trying to set a record or win some money.

Finally, of the leader who had the highest score on the factor of vacillating leadership, one trainee wrote:

If [he] wouldn't lie to the men so much and stop trying to make major . . . this soldier hates his guts for the way he treated me and the rest in basic training.

The last quotation may seem almost as indicative of tyrannical as of vacillating leadership. Indeed, we shall shortly see that there was a high correlation between these two factors.

## LEADERSHIP CLIMATE OF COMPANIES

The computation of factor scores simplifies the data a good deal; instead of there being, for each of 82 leaders, scores on 15 items, there are only the scores on the 3 factors. One more important step remains: to combine the scores for each leader in a company into measures of the *leadership climate*

of that company. At first glance it might appear that one could take a simple average of the leadership factor scores for the leaders in each company. There are two reasons, however, for not doing this. First, the leaders did not all serve the same length of time; some were with their companies for the entire 16 weeks, but others served as little as 2 weeks of the training cycle. Second, the leaders also varied in the extent of their influence on the trainees. In general, C.O.'s and Field First Sergeants had more influence than did leaders in the other two positions. And, of course, the personal qualities of the leaders also made some of them more significant than others.

Fortunately, one of the questions on the leadership questionnaire made it possible to measure the relative influence of the leaders:

- (1) The four men listed on the right side of this paper are all important in the life of a trainee. Place the No. 1 in the column under the name or names of the men who had the MOST influence in your life as a trainee; the No. 2 in the column under the name or names of the men who had LESS influence and the No. 3 in the column under the name or names of the men who had the LEAST influence or NONE AT ALL.

The average score received by each leader on this question can serve as a measure of his perceived relative importance in determining the leadership climate of his company. Incidentally, the Field First Sergeant had the most influence just as often as the C.O., thus bearing out the point made earlier, that there is more to the effects of leadership than rank alone.

It seems obvious that the dimensions of leadership climate should be the same as the three factors of leader behavior, provided that the scores on these factors can be modified to take into account the variations in length of service and in importance to the trainees. A procedure for doing this uses a modified, or *weighted*, average; each leader's factor scores are given more or less weight according to his length of service and his relative importance. Thus, in the "indexes of leadership climate" for the company, a leader who served all 16 weeks would have his factor scores counted twice as heavily as a leader who served only 8 weeks. Similarly, leaders with high "influence scores" would have their three factor scores weighted more heavily in the indexes of leadership climate than would leaders with low "influence scores."

Computing these weighted averages of the factor scores yields three indexes of leadership climate for each company, one for each factor of leadership. For this study, it suffices to condense these indexes into only two values, "high" and "low" (actually, relatively high and relatively low). A further simplification comes from the high correlation between the indexes of tyranny and vacillation. With only one exception, companies high on tyranny were also high on vacillation. With only 12 companies, it was impossible to separate tyranny from vacillation.

TABLE 3. Types of Leadership Climate

| INDEXES OF LEADERSHIP |                            | LEADERSHIP CLIMATES | NUMBER OF COMPANIES |
|-----------------------|----------------------------|---------------------|---------------------|
| Positive              | Tyrannical and Vacillating |                     |                     |
| High                  | High                       | Paternal            | 1                   |
| High                  | Low                        | Persuasive          | 6                   |
| Low                   | High                       | Arbitrary           | 3                   |
| Low                   | Low                        | Weak                | 2                   |
|                       |                            |                     | Total 12            |

When this is done, there are only four different types of leadership climate, corresponding to high and low values on the first two indexes of leadership climate, as shown in Table 3.

The statistical techniques of averaging, correlation, and factor analysis have made it possible to distill these four types of leadership climate from the 189,000 separate ratings of leaders. Simply in the sense of reducing a mass of virtually indigestible data to a set of straightforward types, this is impressive. Data reduction alone was not the point of this research; rather, it was to study the effect of leadership on nonduty behavior. The value of this statistical analysis thus lies in finding how much difference these types of leadership climate make in the patterns of nonduty behavior. The gross differences in rates of different kinds of behavior between leadership climates are seldom larger than 10 percentage points, but they are remarkably consistent. There is space here only to sketch these effects; for further details the reader may consult Selvin (1960, especially Chapters 5 to 7).

#### EFFECTS OF LEADERSHIP ON BEHAVIOR

Because of a change in the behavior questionnaire during the gathering of the data, it was not possible to compare the frequencies of different kinds of nonduty behavior in the "paternal" climate with the rates in the other three climates. The remaining three climates—"persuasive," "weak," and "arbitrary"—can be thought of as spanning the continuum from competent, democratic, and considerate leadership to incompetent, coercive, and unsympathetic leadership. By and large, these differences in type of leadership correspond to the differences in frequencies and patterns of nonduty activities. The "persuasive" climate has the lowest rates on many of the nonduty activities, the "weak" has intermediate levels, and the "arbitrary" has the highest; or, to put it quantitatively, comparing the rates in the three climates with the rates for all trainees taken together, the rates in the "arbitrary" climate

TABLE 4. Proportion Reporting Getting Drunk at Least Once During Basic Training, by Leadership Climate

|                      |     |
|----------------------|-----|
| "Persuasive" climate | 25% |
| "Weak" climate       | 36% |
| "Arbitrary" climate  | 34% |

are higher than the rates for all trainees in 13 activities, the rates in the "weak" climate are higher in 9 and the rates in the "persuasive" climate are higher in 5 activities.

As an illustration of the type of relation found in this study, consider the effect of leadership climate on the incidence of drunkenness. In answer to the question "How many times during basic training did you get really drunk?" the figures shown in Table 4 were obtained. The maximum difference in this table, 11 percentage points between "persuasive" and "weak" climates, is typical of most of the differences between leadership climates in this study. They were usually no larger than 10 percentage points. This may not seem like much of a difference. Does the smallness of this relation mean that leadership has little effect on nonduty activities? Or does leadership have a larger effect, one that somehow does not appear in these figures?

The latter conjecture seems to be correct. The effects of leadership differences bear unequally on different kinds of men, some showing great differences in their rates of specific nonduty activities from one leadership climate to another and others seeming almost immune to differences in leadership. Thus consider the same relation between leadership and drunkenness, but this time examined separately for single and married men (see Table 5).

Compare the two columns with each other and with the figures in the preceding table. Among the single men, leadership climate has only a small effect on rates of drunkenness; the difference between the highest and lowest is only

TABLE 5. Proportion Reporting Getting Drunk at Least Once During Basic Training, by Leadership Climate and Marital Status

|                      | SINGLE<br>MEN | MARRIED<br>MEN |
|----------------------|---------------|----------------|
| "Persuasive" climate | 30%           | 14%            |
| "Weak" climate       | 38%           | 32%            |
| "Arbitrary" climate  | 33%           | 36%            |

8 percentage points. Among the married men the picture is altogether different. The difference between the highest and lowest rates is 22 percentage points, almost three times as much.

Similar findings hold for most of the activities in this study and for the other two individual characteristics on which data were gathered, age and education. The effects of leadership are felt most among the older, married trainees who had not graduated from high school, and they are felt least among the younger, single high-school graduates. Statisticians express relations like these by saying that the leadership climate and individual characteristics *interact* in their effects on behavior; the effects of leadership climate on behavior depend on the background of the trainee, and, correspondingly, the effects of background on behavior vary from one kind of leadership climate to another. In short, the types of leadership climate constructed by the elaborate statistical procedure described in this chapter not only make sense; they also make a difference.

#### OTHER APPLICATIONS OF STATISTICS IN EVALUATING INDIVIDUALS

The method of describing leadership climate by a combination of statistical procedures appears to be applicable to a wide range of situations in which an individual (a leader, a doctor, a patient, or even an inanimate object like a book, picture, musical performance, or other aesthetic object) is rated independently on a number of variables by a group of judges, each of whom is well acquainted with the individuals he is rating. Perhaps the most significant extension of this work would be its application to other kinds of military units, both in training and in combat. Such studies should also examine what this study chose to ignore, the effects of leadership on the performance of assigned duties as well as nonduty behavior.

#### REFERENCES

- Raymond Bernard Cattell. 1952. *Factor Analysis: An Introduction and Manual for the Psychologist and Social Scientist*. New York: Harper. Probably the clearest account of the principles of factor analysis.
- Harry H. Harman. 1960. *Modern Factor Analysis*. Chicago: University of Chicago. This is a very technical, but encyclopedic, book.
- Hanan C. Selvin. 1960. *The Effects of Leadership*. New York: Free Press.