Journal of Personality and Social Psychology

Volume 7, No. 4
Part 2 of 2 Parts

December, 1967
Whole No. 644

MONOGRAPH

TRAIT INFERENCES:

EVALUATIVE AND DESCRIPTIVE ASPECTS[1]

DEAN PEABODY

Department of Psychology and Education, Swarthmore College

The common emphasis on the importance of evaluation in judgment was critically examined in a study of inferences between traits. Sets of traits were selected to remove the usual confounding between evaluative and descriptive aspects of judgment. Subjects made inferences from 90 traits to 40 scales defined by two opposing traits. Results revealed the secondary importance of evaluation. On 70 items where they were directly opposed, the descriptive aspects were always decisive over evaluation. Factor analysis showed that none of the factors was evaluative. General descriptive dimensions suggested by factor analysis could account for the evaluative consistency of inferences in general. Evidence suggested that evaluation is typically based on a descriptive judgment of the degree of extremeness. A model was proposed for the systematic relation of evaluation to descriptive judgment.

We normally make judgments of people using trait terms such as "kind" or "rash." What are the important aspects of such judgments? Some aspects seem of obvious importance: these judgments typically involve us in liking or disliking the person and in evaluating him as relatively good or bad. Many interpretations emphasize the great importance of such affective or evaluative aspects of judgments. The present paper will critically examine some recent forms of this emphasis. These come from two main sources:

1. *The factor analysis of semantic differential judgments.* Osgood, Suci, and Tannenbaum (1957) analyzed the correlations between judgments on a variety of adjective scales. In several analyses the first and largest factor was regularly interpreted as evaluative. It was claimed that evaluation had a "dominant role [p. 38]" as the "first determinant operating in meaningful judgment [p. 46]."

2. *Theories of the consistency of evaluation or affect.* Logically simpler than the factor analysis of many judgments is the emphasis on evaluative or affective consistency between as few as two judgments. The most general examples of this emphasis are such balance theories as Heider (1958), Osgood and Tannenbaum (1955), Rosenberg and Abelson (1960).

[1] This investigation was supported in part by Public Health Service Research Grant MH-06094 from the National Institute of Mental Health. The author's debt to others is even greater than usual. The study was stimulated by discussions and joint pilot work with S. E. Asch. Over several years, many students—unfortunately too many to list here—collaborated in the work. Martin Ostwald and Richard Brandt, colleagues in classics and philosophy, contributed their broader viewpoints. L. R. Goldberg and N. S. Thompson gave important advice. Alan Coller and Ann Severson served as assistants on the project. At various points, the project benefited from the computer facilities at Haverford College, Swarthmore College, the Health Science Institute in Los Angeles, and the University of California, Berkeley.

The general arguments have been made elsewhere that these interpretations exaggerate the importance of evaluation or affect in several ways. A central argument was that there has been a failure to distinguish clearly between evaluative and descriptive aspects of judgment which are normally confounded with each other.

The present paper is an empirical investigation of these issues as they apply to inferences between personality traits. There are several advantages to considering trait inferences. It is with trait terms that we judge others in everyday life. The case has been eloquently put by Norman[2] for the study of trait terms from the natural language, which he describes as

> the set of *all perceptible variations in performance and appearance between persons or within individuals over time and varying situations* that are of *sufficient social significance, of sufficiently widespread occurrence,* and of *sufficient distinctiveness* to have been encoded and retained as a subset of descriptive predicates in the natural language during the course of its development, growth and refinement.

Moreover, as the pioneering work of Asch (1946) made clear, judgments of different traits are generally not independent but highly related. To examine the inferences between trait terms thus becomes a way of getting at what Cronbach (1955) called the "implicit personality theories" that people use in their everyday judgments.

At the same time trait inferences are representative of both forms of the claims for the importance of evaluation. (*a*) Osgood (1962) described factor analyses of trait inferences by himself and Ware. He claimed to find "such large evaluative factors (58% and 69% of total variance, respectively) that little else could be determined . . . [p. 25]." (*b*) It is logically simpler, however, to begin with the emphasis on evaluative consistency within separate inferences. An example is the work of Podell (1961). His subjects were given three favorable or three unfavorable traits as characterizing a person. They then selected other traits as also likely for the person. The selected traits were very similar to the given traits in their evaluative ratings. Podell (1961) concluded that "the judgments were demonstrated to have been determined by this [evaluative] factor," although "subjects did not deliberately judge on this basis [p. 596]." Moreover, Podell gave a similar reinterpretation to the earlier findings of Asch (1946) of very different impressions for persons described as "warm" or "cold"; these findings were now "explained in terms of the substantial difference in favorableness between the two terms [Podell, 1961, p. 596]."

Podell's (1961) subjects inferred from certain traits to other traits. Other studies (Anderson, 1962; Fishbein & Hunter, 1964; Podell & Amster, 1966) go further, and have inferences made directly onto such purely evaluative scales as *good-bad.* These studies involve two different issues: (*a*) they share an implicit assumption that an adequate analysis need consider only evaluation; (*b*) they go on to disagree about the particular formula for combining evaluations of several traits. The two issues can, however, be separated. Asch (1946) originally raised the issue of combining several traits, arguing against combination based on some form of addition. On the other side of this issue, Bruner, Shapiro, and Tagiuri (1958) used an analysis implying additive combination but not necessarily on an evaluative dimension. In contrast, the present study puts aside the question of combination in order to concentrate on the prior issue: can evaluation account adequately for inferences from a single trait?

Plan of the study. The problem is that a judgment is commonly at the same time both an estimate of the factual situation and an evaluation. For example, consider such contrasts as "kind-cruel" or "cautious-rash." The judgment that a person or action is "kind" or "rash" combines a descriptive aspect (e.g., that the action helps others, or involves very large risks) and an evaluative aspect (e.g., that the action is desirable or undesirable). In a single trait term (and its opposite) the two aspects are confounded since they always combine in the same way (e.g., helping as desirable, hurting as undesirable). Further, two related traits may be similar in both descriptive and evaluative aspects. If an inference is made between two such traits, is it based on the evaluative or the descriptive similarity? If Podell's (1961) subjects inferred that a person who was "considerate"

[2] Warren Norman, unpublished manuscript.

TABLE 1

SCHEME FOR UNCONFOUNDING EVALUATION AND A DESCRIPTIVE ATTRIBUTE

	Descriptive attribute	
	X	Un-X
Evaluation		
+	Term 1	Term 2
−	Term 3	Term 4

was also likely to be "kind," was this simply because both terms were favorable, or was it because they share a common descriptive quality like helpfulness, which was evaluated favorably?

There are then two opposing interpretations, and according to either, evaluative similarity would be a typical result of trait inferences: (*a*) For theories of evaluative consistency, represented for trait inferences by Podell (1961), the evaluative similarity between traits would be of primary importance[3]; (*b*) for the alternative interpretation, the descriptive similarity between traits would be of primary importance. Evaluative similarity would tend to occur only because it is the typical by-product of descriptive similarity.

To test these two interpretations, what was needed was some method of removing the usual confounding between evaluative and descriptive similarity. The method finally devised was a direct one: sets of traits were selected to separate similarity in evaluation from that on some descriptive attribute. The scheme is represented in Table 1. Once the confounding between evaluative and descriptive aspects was removed, it became possible to see what was of greater importance for inferences between the traits.

METHOD

Selection of Sets of Trait Terms

It was not difficult to think of instances of trait terms with the needed relations. Since, however, it was intended to have a fairly comprehensive coverage of trait terms in English, various sources were examined for possible terms. As general sources, there were dictionaries of synonyms and antonyms, Roget's *Thesaurus*, and a condensation of the latter by Osgood et al. (1957) listing 289 pairs of adjectives. Specific to personality traits was the listing of Allport and Odbert (1936), which in turn had been condensed by Cattell (1946) into 171 groups of terms.

The selection of trait terms was carried out in several steps, using students in an advanced seminar as an advisory panel. In the first step, most trait terms that were not obviously uncommon were classified, according to the scheme of Table 1, into tentative sets having roughly the same attribute. Nearly 700 traits were classified into about 45 tentative sets. These tentative sets were of three types:

Sets of four. The majority of the sets included terms for all four combinations represented in Table 1. As an illustration, consider the set that was eventually represented by the terms "cautious," "timid," "bold," and "rash." At this early stage, the tentative set included in addition near-synonyms of these terms such as "prudent" and "reckless," and also related terms such as "cowardly" and "brave" that were clearly distinct, but could not be built up into a separate set of their own.

Sets of three. Most of the remaining sets included three of the possible combinations. Typically, what was missing was a single term in English to represent one of the favorable combinations. This type of set was common among traits involving ability; for example, while there are terms for a negative form of intelligence such as "crafty," there seems to be no single term for a positive form of stupidity.

Sets of two. In a relatively few sets, there were terms only for two opposite combinations (represented diagnonally in Table 1). Examples would be such contrasts as "honest" versus "dishonest" or "fair" versus "unfair."[4] Thus, there seems to be no

[3] "Consistency" implies an inference from a trait (*a*) to a similar trait; (*b*) away from a dissimilar trait. The analysis here generally emphasizes the former, and "similarity" can be used as a more specific substitute for "consistency."

[4] Further examples are suggested by the work on the "interpersonal dimension of personality" (e.g., Leary, 1957) as measured by the Interpersonal Check List (Laforge & Suczek, 1955). This approach represents interpersonal traits of personality in a circular schema, which is summarized by two axes of dominance-submission and love-hate. This analysis resembles the present one in considering evaluation separately from (nonevaluative) personality characteristics, and as a function of their intensity.

For the love-hate dimension, however, the separation is not successful by the present criterion: there is a lack of single terms in English for undesirable forms of love or desirable forms of hate. For example, at the level of greatest desirability all five of the items on the Interpersonal Check List that are scored heavily for love are single words—(i.e., "friendly," "considerate," "cooperative," "helpful," "appreciative"), whereas the five comparable items scored heavily for hate are all phrases—(i.e., "Can be frank and honest," "Can be strict if necessary," "Can complain if necessary," "Able to take care of self," "Able to doubt others"). In general, traits representing "love-hate" cannot have their con-

single term to represent either a negative form of honesty or a positive form of dishonesty. Although honesty is logically distinguished from evaluation in having a specific factual content, the confounding between them cannot be removed by the present method. Since such terms cannot contribute to the further analysis, which requires removing the confounding, they were eliminated at this stage. To this extent, the analysis does not represent the entire range of possible traits.

However, there remained a large proportion of trait terms included in the tentative sets of four and three. These sets were now used for the second step in selection. Within each set, two terms were selected to represent each combination (except where only one term was available). The criteria were to choose terms that were relatively (*a*) common and general, (*b*) close to the common attribute of the set, (*c*) likely to receive clear-cut evaluative ratings. The chosen terms were combined with a few neutral terms in a list of 289 terms. This list was given preliminary evaluative ratings by 20 student judges on 7-point scales with the usual semantic differential procedure, 10 judges each using the scales "favorable-unfavorable" and "desirable-undesirable." (Twenty additional judges later made similar judgments for the terms finally selected; the evaluative ratings reported below are for the entire 40 judges.) The preliminary ratings from the two scales were combined, and the *t* tests were made to see which terms were significantly positive or negative in evaluation.

The evaluative ratings were then used in the final selection of sets and specific terms to represent them. Questionable sets were eliminated, using as criteria that (*a*) a common attribute should be central to each of the terms in a set; (*b*) related sets should be kept separate if there were distinguishable attributes represented by terms for at least three of the combinations, but combined otherwise; (*c*) positive and negative terms should differ significantly in evaluation. Indeed, in the final selection, each separate term was significantly positive or negative, except for one term ("impulsive") with only borderline significance ($p = .06$). Terms for 15 sets of four and 10 sets of three were finally selected. These trait terms are presented with their evaluative ratings in Table 2, using the same format as Table 1. The traits in Table 2 are roughly grouped according to their general content (i.e., temperament, social, ability, and ideas).

Although the traits in each set were intended to share a common attribute, inspection of Table 2 indicates that they do so with varying success. For this initial investigation it seemed preferable to retain breadth of coverage rather than use only the best sets of terms.

The selection of terms necessarily involved difficult judgments at several points. Some of the decisions were certainly arguable; for example, the specific

TABLE 2

SETS OF TRAIT-TERMS SELECTED AND THEIR EVALUATIVE RATINGS

Set no.				
	Sets of four			
	Temperament			
1	+ .9	Cautious	+1.1	Bold
	−1.1	Timid	−1.2	Rash
2	+1.7	Self-Controlled	+1.1	Uninhibited
	−1.4	Inhibited	− .3*	Impulsive
3	+1.3	Serious	+1.5	Gay
	−1.6	Grim	−1.2	Frivolous
4	+2.0	Alert	+1.8	Relaxed
	−1.1	Tense	−1.7	Lethargic
5	+ .8	Committed	+2.5	Open-Minded
	−2.4	Fanatical	− .8	Noncommital
6	+1.3	Steady	+1.6	Flexible
	−2.1	Inflexible	−1.5	Vacillating
7	+2.0	Modest	+1.3	Confident
	−1.1	Self-Disparaging	−2.0	Conceited
	Social			
8	+ .9	Thrifty	+1.8	Generous
	−2.0	Stingy	− .8	Extravagant
9	+ .5**	Skeptical	+1.1	Trusting
	−1.4	Distrustful	−1.4	Gullible
10	+1.3	Selective	+2.5	Tolerant
	− .5**	Choosy	−1.4	Undiscriminating
11	+1.3	Firm	+ .9	Lenient
	−1.4	Severe	− .9	Lax
12	+1.3	Discreet	+1.8	Frank
	−1.2	Secretive	−1.4	Indiscreet
13	+2.0	Individualistic	+1.6	Cooperative
	−2.0	Uncooperative	−1.6	Conforming
	Ideas and ability			
14	+ .9	Pragmatic	+1.5	Idealistic
	− .6**	Opportunistic	−1.2	Unrealistic
15	+1.6	Cultivated	+2.1	Natural
	−2.2	Artificial	− .7	Naive
	Sets of three			
	Temperament			
16	+1.8	Thorough		
	−1.5	Fussy	−1.4	Careless
17	+1.3	Moral		
	−1.6	Self-Righteous	−1.6	Immoral
18	+1.8	Curious		
	−1.7	Nosy	−1.5	Uninquisitive
	Social			
19	+1.0	Forceful		
	−2.0	Domineering	−1.4	Submissive
20	+1.7	Peaceful		
	−1.2	Passive	−2.0	Belligerent
21	+1.6	Polite		
	−1.5	Ingratiating	−2.2	Rude
	Ideas and ability			
22	+2.6	Intelligent		
	− .7**	Crafty	−1.7	Stupid
23	+1.9	Foresighted		
	−1.3	Scheming	−1.5	Short-Sighted
24	+1.6	Meditative		
	− .9	Brooding	−1.6	Unmeditative
25	+1.8	Witty		
	− .8	Sarcastic	−2.1	Humorless

Note.—The evaluative ratings are means for 40 judges. The scales "favorable-unfavorable" and "desirable-undesirable" were used by 20 judges each, and scored from +3 to −3. The means are significantly different from zero with $p < .001$ for 85 of the 90 terms; for the remaining five terms the significance level is indicated.

* $p = .06$.

** $p < .01$.

founding with evaluation removed using the present method, and would correspond to "sets of two." (In contrast, the dominance-submission axis has an analogue in Factor II of the analysis below.)

term chosen (e.g., "rash" or "reckless"); whether an attribute like dominance-submission (Set Number 19) should be kept separate from others such as belligerence (Set Number 20) and cooperation (Set Number 13). On the other hand, it was easy to show that the selection was not idiosyncratic.[5] Moreover, the general adequacy of the selection was subject to empirical confirmation from the results of inferences between the traits: only if the terms in a set were actually related in some important respect that contrasts with evaluation is there any reason why inferences should systematically go contrary to evaluative similarity.[6]

Collection of Inference Data

The removal of the usual confounding now made it possible to examine trait inferences where evaluative and descriptive similarity were opposed to each other. In the critical inferences, subjects would be given a single trait (the "antecedent term") and a choice between two alternative traits. In relation to the antecedent term (e.g., CAUTIOUS), the two alternatives would be (*a*) similar in evaluation and dissimilar on the descriptive attribute (e.g., *bold*), and (*b*) dissimilar in evaluation and similar on the descriptive attribute (e.g., *timid*). Thus, the two alternatives would be opposites of each other, as represented diagonally in Tables 1 and 2. Following Osgood, these pairs were used to define the ends of seven-step scales.

[5] For example, independent judges made the same classification of the selected traits. To avoid dealing with too many terms at once, the 15 sets of four were divided randomly into (*a*) five combinations of three sets, or (*b*) three combinations of five sets. Within each combination, the positive and negative traits were arranged alphabetically. Two student judges were asked to classify them according to the format of Table 1. Their classification agreed perfectly with that of Table 2.

[6] On this point, compare the work of Steiner (1954). He paired one trait term with two others, one pair being more similar in evaluation, and asked subjects to choose which pair was more likely to go together. He never makes it explicit that the second pair was more similar in some other (i.e., descriptive) respect, but unless this was so, it is hard to see why any subjects should choose it. In fact, this may be the explanation for the large differences he found between different items: for some items, there may be a descriptive basis for choosing against evaluative similarity and most subjects do so —for example, 23/26 put "quick-tempered" (−1.76) with "passionate" (+0.47) rather than with "lazy" (−2.52). For other items, descriptive relations may not oppose evaluative similarity—for example, 24/26 subjects put "religious" (+1.92) with "practical" (+1.69) rather than "sophisticated" (+0.22). On this interpretation, Steiner's results are in agreement with the present study, where descriptive and evaluative similarity were explicitly and deliberately separated and opposed.

In general, each inference item was from a single antecedent term to a scale defined by two opposite traits. Altogether, each of the 90 selected traits (from 15 sets of four plus 10 sets of three) was used as an antecedent term. There were 40 different scales (i.e., two scales from each of the 15 sets of four, one scale from each of the 10 sets of three). In all, therefore, there were inferences from 90 antecedent terms to 40 scales—a total of 3,600 items.

Osgood et al. (1957) had subjects respond to 1,000 items or more, but this many items seemed likely to produce considerable psychological satiation. Some experimentation was done on this problem in the present study: approximately half the subjects were given 240 items and the others 360 items. Some of the latter subjects complained that they were responding randomly at the end, but analysis of the results showed that this was not the case. Altogether, there were 240 subjects, students at Swarthmore College largely from the introductory psychology course. For each item, there were 20 subjects and the entire analysis is based on their item means.

The administrative procedure generally resembled that for the semantic differential (Osgood et al., 1957, pp. 82–84), but followed the adaptation used by Osgood and Ware (Osgood, 1962, p. 25) in asking for the likelihood of inference rather than the degree of meaning. The only major difference from the Osgood and Ware procedure was that the antecedent term was presented without a contrasting opposite. On each item, the antecedent term was printed in capitals above the middle of the scale. The instructions said: "assume a person with the characteristic given in capital letters. On the scale immediately beneath, you are to judge how likely it is that this person has one or the other of the traits given by the scale." A sample item would be:

CAUTIOUS

bold ——:——:——:——:——:——:—— timid

The arrangement of items in booklets was controlled and counterbalanced in several ways. The 40 scales were ordered on two pages of 20, so as to separate related scales. The favorable end of successive scales alternated between right and left positions. Half the subjects received alternate forms with the order of the 40 scales and their left-right arrangement reversed. There were 90 different sequences, systematically rotating each of the 90 antecedent terms across the 40 scales. The alternate forms used the exact reverse of each sequence. An individual subject received either 6 or 9 of these sequences (240 or 360 items).

Although one objective had been to select terms of relatively common usage, the nature of the design required less common terms (e.g., if politeness is normally positive, its negative form will have relatively infrequent usage). After one subject asked specifically about the meaning of "ingratiating," supplementary instructions were added that gave a dictionary definition of this word and suggested the use of an available dictionary in any further cases of doubt. Despite the high verbal aptitude of

the subjects, there were doubtless still some cases where a term was not adequately understood.

ANALYSIS OF INDIVIDUAL ITEMS

Each response was made on a seven-step scale, and was scored from +3 to −3 between the favorable and unfavorable terms defining the scale. A positive score always represented the evaluatively favorable direction. For each item, the mean score for the 20 subjects was calculated, and a t test was made as to whether this mean differed significantly from zero. The basic question for each item was to compare the sign of the item mean with the evaluative sign of the antecedent term. Depending on whether these two signs were the same or the opposite, the item was counted as representing evaluative similarity or dissimiliarity.

Unselected Inferences

Consider first trait inferences in general. Of the 3,600 items, the great majority (3,450 items) were inferences between sets, from an antecedent term in one set to a scale defined by two terms from another set. The design was such that these items should represent a broad range of traits, but it did not establish in advance any descriptive relations between them. Accordingly, it was a striking preliminary finding that subjects showed a very high degree of consensus as to the direction of inferences: 46% of all the item means were significantly different from zero at the .001 level, and 69% at the .05 level. Far from restricting inferences only to closely related terms, subjects demonstrated a very extensive common network of inferences.

For such unselected inferences, it was argued earlier that evaluative and descriptive similarity are typically confounded: inferences generally tend to be between traits that are similar descriptively, and typically also receive similar evaluations. Evaluative similarity would be typically expected by either interpretation, whether interpreted directly or as an indirect by-product of descriptive similarity.

As expected, evaluative similarity was the typical result, occurring on 67% of the 3,450 items (and 79% of the 1,603 items that were significant at the .001 level). These are impressive majorities of evaluative similarity. Such frequencies of evaluative similarity for cognitions about traits almost certainly exceed those for actual traits of people, an illustration of the classic finding of the "halo effect."

Critical Inferences Within Sets

In contrast, consider the critical items where the inference to a scale is from another trait in the same set (e.g., from CAUTIOUS to the scale *bold-timid*).[7] These were the items for which the terms were selected so that evaluative and descriptive similarity would oppose each other. The results were clear-cut. There were 70 such critical items (i.e., 4 from each set of four and 1 from each set of three). On all 70 item means, subjects inferred toward descriptive similarity and away from evaluative similarity, significantly so on 66 of them. Inferences from postive traits were toward negative traits, and inferences from negative traits were toward positive traits. When they were confronted directly, descriptive similarity was decisive over evaluative similarity.

A common reaction to these results is to feel that they are (*a*) obvious and (*b*) not really crucial for the usual claims for evaluation. For example, it might be objected that Osgood and others admit the existence of nonevaluative aspects of meaning. They might grant the importance of very specific nonevaluative linkages between words that are nearly synonyms rather than antonyms, as in inferring from "cautious" to "timid" rather than "bold." Such special cases would be exceptions to the general importance of evaluation. The present analysis—in treating these special cases as of critical significance—would be attacking a straw man.

The answer to this objection has at least two parts:

1. The interpretations emphasizing evaluation have been protected by their vagueness. In particular, they have never dealt with the problem of confounding between evaluative and nonevaluative aspects of judgment. Al-

[7] For the sake of continuity, examples are usually taken from this first set of traits. Where appropriate, the convention of Osgood is followed here, using small capitals for the antecedent terms and italics for those defining the scale.

though admitting that nonevaluative aspects exist, these interpretations do not state clearly when—if ever—they would be decisive over evaluation. It is true that on the other hand, these interpretations do not state clearly that evaluation always tends to be decisive—although this is the impression that is given (for example, by such general statements as those quoted at the beginning, and by the formulas dealing only with evaluation). In view of this lack of clarity, it seemed an appropriate first step to confront evaluative and descriptive similarity with brutal directness. The results then showed that in every case descriptive similarity was decisive. To call these results obvious is to admit that the unlimited claims for evaluation are obviously wrong, once they are put more clearly.

2. The second part of the answer is that the critical inferences are indeed special cases, and need to be supplemented by further analysis. The results presented thus far do not disprove the objection that specific descriptive similarity may only be an exception to the general importance of evaluation. This objection is dealt with in the remaining sections, which consider evidence that general descriptive similarity is also decisive over evaluation.

At this point, it may sharpen the issues to state as clearly and strongly as possible the alternative interpretation: Although evaluation is normally confounded with descriptive aspects of judgment, it tends always to be secondary to them. (*a*) Some descriptive aspects can be separated from evaluation and then shown to be of greater importance. Such aspects typically include similarity of descriptive direction on (1) a specific attribute, or (2) a more general dimension. The design of the present study established in advance only some relatively specific descriptive attributes; the previous section showed that they accounted for the inferences within sets. The evaluative similarity of inferences in general (e.g., between traits in different sets) was left unaccounted for. The next section uses factor analysis to suggest what any general descriptive dimensions might be. It then examines evidence that similarity on such a dimension can account for the evaluative similarity of inferences in general; (*b*) other descriptive aspects may be the basis of evaluation itself. Such aspects must necessarily remain confounded with evaluation. A subsequent section examines evidence that, at least for traits, the basis of evaluation is typically a descriptive judgment of the degree of extremeness, with positive and negative traits being typically moderate and extreme.

FACTOR ANALYSIS OF TRAIT INFERENCES

To have begun with the analysis of separate inferences items was appropriate since it is crucial that an interpretation of trait inferences account for them. There were, however, several reasons for proceeding to the factor analysis of the correlations between them. (*a*) Much of the current emphasis on evaluation was inspired by claims that large evaluative factors were repeatedly found in the factor analysis of semantic differential ratings. However, the interpretation of these factors is ambiguous, since they may involve the confounding of evaluative and descriptive aspects. By removing the confounding, the present design makes possible a clear interpretation of whether any factors are evaluative or descriptive. (*b*) As was just pointed out, factor analysis might also be of use in extending the analysis of separate items. The factors may suggest what any general descriptive dimensions might be. One can then examine whether similarity on such a descriptive dimension could account for the evaluative similarity of inferences between different sets. (*c*) Finally, any factors are of interest in their own right, as the result of an analysis of a broad range of trait judgments. The factors suggest the main dimensions of the "implicit personality theories" (Cronbach, 1955) that people use in their cognitions of each other in everyday life. In particular, the factors that summarize these cognitions can be compared with those from actual measures of personality. Since cognitions tend to regularize reality, they may be expected to yield a simpler factorial summary.

Unlike most factor analyses, the present one was not based on correlations across individual subjects. The basic data for the analysis consisted of the item means (each representing 20 subjects) for inferences from the 90 antecedent terms to the 40 scales. This

matrix of item means was analyzed in both possible ways: (*a*) an analysis of the 90 antecedent terms correlated across the scales; (*b*) an inverse analysis of the 40 scales correlated across the antecedent terms.

In each case, the correlations were factored by the principal axis method. The factors were then rotated using the varimax procedure. These objective procedures do not yet completely remove the need for decisions. For example, the marginal decision whether or not to recognize a small factor may have an important influence on the subsequent varimax rotation. This problem was met by comparing several alternative procedures; for example, the use of commonality estimates or of unity in the diagonals, which—with these large numbers of variables—can make only a small difference. The relatively conservative procedure was finally followed of recognizing only factors with Eigenvalues that were greater than 1.0 using the commonality estimates.

The method yielded seven factors for the analysis of antecedent terms, and four factors for the analysis of scales. Among the former seven, there was a precipitous drop from the first three rotated factors (each of which represented 25% or more of the common variance), to the last four factors (each of which represented only around 4%). The three major factors will be given primary attention here. As proportions of total variance, the three major factors together represent 75%, as compared with 90% for all seven factors, and 84% for the four factors in the analysis of scales. These proportions are large: compared with typical analysis of actual trait measures, cognitions about traits yield a much simpler and more powerful summary.

The varimax rotation procedure—an objective version of Thurstone's earlier principle of simple structure—is intended to yield rotated factors that are more interpretable than the unrotated ones. On the other hand, Thurstone's principle was designed to break up any general factor, and it might be claimed that evaluation is likely to be a general factor that would be destroyed by rotation. However, in this case, the unrotated first factor yielded by the principal axis method should represent a general factor to the extent that there is one. Accordingly, for each analysis results are reported below for the first unrotated factor and for the major rotated factors.

Analysis of Antecedent Terms

Table 3 presents factor loadings for the 90 antecedent terms. The main part of the table presents loadings on the first unrotated factor (I_U) and the three major rotated factors (I, I, and III). The end of the table lists the larger loadings on the four minor factors. In the main part of the table, the antecedent terms within each set have been ordered so as to alternate direction on the descriptive attribute. If a factor represented these descriptive directions, the larger loadings should therefore alternate. If a factor were evaluative, the loadings should instead vary with the evaluative sign of each trait.

The central result is clear-cut; none of these large factors is evaluative. Instead, the substantial loadings systematically alternate according to the descriptive direction. Thus, removing the usual confounding between evaluation and descriptive aspects removes the possibility of interpreting any factor as evaluative.

The first unrotated factor (I_U) represents 42% of the total variance and is quite pervasive (although it is not technically a "general" factor where all loadings tend to be positive). The large loadings are too numerous to list here: considering only sets where all the traits have loadings larger than .60 would include sets 2, 3, 8, 9, 10, 11, 16, and 24. It is hard to find a name for such a broadly pervasive factor. It represents a very general contrast that might be called "tightness versus looseness of control." More imaginative titles might be chosen, such as "anal character," or "the Protestant ethic." Related factors, often called "introversion versus extraversion," have frequently appeared in analyses of actual trait measures.

As anticipated, this unrotated factor was partly broken up by rotation, largely into the Rotated Factors I and III. The relation between the three factors is shown by the index of factorial similarity (e.g., Harman, 1960, p. 251): .90 between Factors I_U and

TABLE 3

FACTOR ANALYSIS OF ANTECEDENT TERMS

Set no.	Trait no.	Trait	Evaluative sign	Unrotated 1st factor I_U	Major rotated factors			
					I[a]	II[a]	III . . .	h[a b]
		Sets of four						
		Temperament						
1	1	Cautious	+	76	86	−28	25	90
	2	Bold	+	−10	−40	81	10	92
	3	Timid	−	07	33	−82	−06	88
	4	Rash	−	−44	−64	67	−15	94
2	5	Self-Controlled	+	70	72	−35	40	88
	6	Uninhibited	+	−71	−65	49	−49	97
	7	Inhibited	−	68	62	−49	45	95
	8	Impulsive	−	−63	−67	52	−37	91
3	9	Serious	+	85	68	−06	61	94
	10	Gay	+	−88	−69	12	−64	94
	11	Grim	−	91	59	07	73	92
	12	Frivolous	−	−83	−71	19	−57	91
4	13	Alert	+	60	68	50	−06	90
	14	Relaxed	+	−86	−55	−21	−59	95
	15	Tense	−	86	64	22	49	91
	16	Lethargic	−	−53	−28	−71	−24	91
5	17	Committed	+	57	03	23	88	88
	18	Open-Minded	+	−51	06	06	−80	88
	19	Fanatical	−	44	−13	57	70	88
	20	Noncommittal	−	−03	46	−46	−54	88
6	21	Steady	+	59	37	−33	71	93
	22	Flexible	+	−71	−17	−23	−84	93
	23	Inflexible	−	75	19	27	90	96
	24	Vacillating	−	−51	−10	−07	−79	89
7	25	Modest	+	−02	25	−88	−02	89
	26	Confident	+	−15	−43	50	48	88
	27	Self-Disparaging	−	26	45	−48	−02	79
	28	Conceited	−	37	−07	76	41	87
		Social						
8	29	Thrifty	+	92	73	01	58	90
	30	Generous	+	−91	−67	−29	−50	92
	31	Stingy	−	93	62	27	62	94
	32	Extravagant	−	−79	−76	33	−47	94
9	33	Skeptical	+	83	82	39	16	92
	34	Trusting	+	−82	−65	−54	−24	94
	35	Distrustful	−	89	76	30	34	92
	36	Gullible	−	−86	−73	−46	−30	94
10	37	Selective	+	91	77	30	39	89
	38	Tolerant	+	−71	−18	−53	−68	89
	39	Choosy	−	93	68	37	49	90
	40	Undiscriminating	−	−92	−84	−24	−33	92
11	41	Firm	+	78	31	29	85	94
	42	Lenient	+	−88	−44	−47	−71	93
	43	Severe	−	86	38	34	82	94
	44	Lax	−	−92	−61	−29	−65	93

Note.—Additional factors (loadings .30 or greater): Factor IV (4.7% of common variance): artificial .85, ingratiating .67, cultivated .58, conforming .40, crafty .33; lethargic −.30, natural −.43, frank −.54; Factor V[a] (3.9% of common variance): frank .47, open-minded .44, cooperative .43, curious .37, alert .32, uncooperative −.54; Factor VI (3.9% of common variance): idealistic .80, unrealistic .62; crafty −.30; immoral −.30, opportunistic −.35, ingratiating −.45, pragmatic −.53; Factor VII (3.9% of common variance): confident .62, relaxed .44, steady .34, conceited .30; tense −.41, brooding −.42, vacillating −.45, self-disparaging −.58.

[a] This factor has been reflected for convenience in interpretation.

[b] The commonalities represent all seven factors, not only the first three.

Table 3—Continued

Set no.	Trait no.	Trait	Evaluative sign	Unrotated 1st factor I_U	Major rotated factors			
					I_a	II_a	III . . .	h^{2b}
12	45	Discreet	+	68	77	−35	27	89
	46	Frank	+	−29	−40	46	00	88
	47	Secretive	−	87	81	00	35	88
	48	Indiscreet	−	−67	−73	50	−37	95
13	49	Individualistic	+	21	21	83	−15	85
	50	Cooperative	+	−58	−21	−74	−37	93
	51	Uncooperative	−	62	25	63	43	94
	52	Conforming	−	−08	−11	−78	21	90
		Ideas and ability						
14	53	Pragmatic	+	74	64	15	37	92
	54	Idealistic	+	03	−17	18	32	93
	55	Opportunistic	−	46	42	67	−04	86
	56	Unrealistic	−	−37	−61	24	06	88
15	57	Cultivated	+	50	68	−10	−05	86
	58	Natural	+	−77	−59	06	−43	85
	59	Artificial	−	44	29	02	19	86
	60	Naive	−	−75	−69	−42	−17	81
		Sets of three Temperament						
16	61	Thorough	+	86	69	08	56	85
	62	Careless	−	−89	−76	12	−54	93
	63	Fussy	−	85	49	38	60	86
17	64	Moral	+	57	23	−38	78	90
	65	Immoral	−	−47	−29	46	−63	80
	66	Self-Righteous	−	75	21	31	85	89
18	67	Curious	+	−16	14	58	−60	93
	68	Uninquisitive	−	−42	−53	−67	20	92
	69	Nosy	−	30	08	78	06	76
		Social						
19	70	Forceful	+	62	19	74	57	95
	71	Submissive	−	−47	−19	−88	−25	95
	72	Domineering	−	66	16	63	70	94
20	73	Peaceful	+	−48	−15	−86	−22	92
	74	Belligerent	−	46	01	84	45	93
	75	Passive	−	−37	−02	−92	−24	94
21	76	Polite	+	06	33	−78	−03	91
	77	Rude	−	15	−20	90	16	95
	78	Ingratiating	−	−11	04	−16	−31	85
		Ideas and ability						
22	79	Intelligent	+	44	78	29	−31	87
	80	Stupid	−	−52	−85	−20	−28	90
	81	Crafty	−	74	73	37	10	88
23	82	Foresighted	+	72	89	05	10	89
	83	Short-Sighted	−	−57	−87	25	03	84
	84	Scheming	−	80	69	47	23	89
24	85	Meditative	+	68	86	−23	15	90
	86	Unmeditative	−	−68	−90	10	−01	89
	87	Brooding	−	75	68	05	34	89
25	88	Witty	+	−27	−07	64	−62	89
	89	Humorless	−	81	49	−14	78	91
	90	Sarcastic	−	49	27	88	14	86
		% Total variance		(42.0)	29.6	23.6	22.2 . . .	90.1
		% Common variance			32.8	26.2	24.6 . . .	100

I; .32 between Factors I_U and III; .54 between Factors I and III. In comparison to the unrotated factor (I_U), the first rotated factor (I) seems to represent a somewhat more limited form of control which might be called "ego-control versus "impulse-expression." It has high loadings on most of the same traits, and higher ones for some traits of the intellect. This factor can be illustrated by sets where all four traits have loadings greater than .60. Balancing positive loadings against negative ones (and with evaluative signs in parentheses), there is SELF-CONTROLLED (+) and INHIBITED (−) versus UNINHIBITED (+) and IMPULSIVE (−); THRIFTY (+) and STINGY (−) versus GENEROUS (+) and EXTRAVAGANT (−); SKEPTICAL (+) and DISTRUSTFUL (−) versus TRUSTING (+) and GULLIBLE (−). There are also traits of the intellect from sets of three: INTELLIGENT (+) and CRAFTY (−) versus STUPID (−); FORESIGHTED (+) and SCHEMING (−) versus SHORT-SIGHTED (−); MEDITATIVE (+) and BROODING (−) versus UNMEDITATIVE (−).

Rotated Factor III—the other main derivative of the unrotated first factor—might be called "decidedness versus undecidedness." Illustrative sets with at least three loadings greater than .60 would include: COMMITTED (+) and FANATICAL (−) versus OPEN-MINDED (+) and NONCOMMITAL (−); STEADY (+) and INFLEXIBLE (−) versus FLEXIBILE (+) and VACILLATING (−); FIRM (+) and SEVERE (−) versus LENIENT (+) and LAX (−); MORAL (+) and SELF-RIGHTEOUS (−) versus IMMORAL (−).

Factor II is the remaining major rotated factor. This factor might be called "dominance versus submission," or better "self-assertiveness versus unassertiveness" to suggest the inclusion of traits not directly social such as BOLD (+) versus TIMID (−) and CONCEITED (−) versus MODEST (+), in addition to directly interpersonal traits, such as the following: INDIVIDUALISTIC (+) and UNCOOPERATIVE (−) versus COOPERATIVE (+) and CONFORMING (−); FORCEFUL (+) and DOMINEERING (−) versus SUBMISSIVE (−); BELLIGERENT (−) versus PEACEFUL (+) and PASSIVE (−). Similar factors have appeared in other analyses, such as the Dominance-Submission axis of the Interpersonal Check List.

Analysis of Scales

Table 4 presents results for the inverse analysis of the 40 scales. The table lists loadings for the unrotated first factor and the four rotated factors. The first trait listed for each scale is the evaluatively favorable end, which was scored as positive. Hence, if a factor were evaluative, all scales should tend to have similar loadings. If a factor were descriptive, the two scales from the same set should be dissimilar for any substantial loadings.

It is evident that none of the factors are evaluative. Instead, where one scale from a set has a substantial positive loading, the other has a dissimilar loading. These contrasting loadings are of two kinds: (*a*) For the unrotated first factor (i_u), the contrasting loadings tend to be negative ones of equivalent magnitude. A similar balanced "bipolar" pattern was characteristic of all the factors in the analysis of antecedent terms; (*b*) for the rotated scale factors (i–iv), the contrasting loadings tend to be small and typically negative. Such an unbalanced pattern is the likely outcome of the rotation procedure. The rotation generally assigns scales from the same set (whose average correlation is .44 in the descriptive direction) to different factors. Its failure to do the same earlier to antecedent terms from the same set reflects their still higher correlations (averaging .70 in the descriptive direction). A difference of degree in the correlations of the two analyses became a considerable difference in the rotated factors.

Specific interpretation of the scale factors can make use of the unusual relation between the two analyses: the same traits that defined the two ends of each scale were also used separately as antecedent terms. As a result, it is possible to make an unusual comparison between the two analyses, using a modified version of the index of factorial similarity.[8]

[8] This index is generally designed to compare factor loadings using the same variables with different subjects. To compare the present two analyses, it was modified to deal with variables that are different but related. Consider a factor loading for a scale—for

TABLE 4

FACTOR ANALYSIS OF SCALES

Set no.	Scale no.	Scale	Unrotated 1st factor i_u	Rotated factors				h^2
				i	ii	iii	iv	
		Sets of four						
		Temperament						
1	1	Cautious–Rash	72	84	−10	−44	06	91
	2	Bold–Timid	03	−15	−16	93	10	92
2	3	Self-Controlled–Impulsive	73	87	01	−25	25	89
	4	Uninhibited–Inhibited	−65	−46	45	69	−25	96
3	5	Serious–Frivolous	85	80	−22	−03	45	89
	6	Gay–Grim	−77	−38	73	28	−28	83
4	7	Alert–Lethargic	54	58	−11	64	−07	77
	8	Relaxed–Tense	−63	−18	85	13	01	77
5	9	Committed–Noncommittal	36	02	−32	44	71	81
	10	Open-Minded–Fanatical	−24	37	86	04	−17	91
6	11	Steady–Vacillating	58	57	02	23	64	78
	12	Flexible–Inflexible	−53	06	84	01	−45	92
7	13	Modest–Conceited	−11	36	68	−50	07	85
	14	Confident–Self-Disparaging	−04	04	23	73	20	63
		Social						
8	15	Thrifty–Extravagant	89	82	−32	−18	29	89
	16	Generous–Stingy	−71	−26	90	12	04	89
9	17	Skeptical–Gullible	85	73	−50	28	−16	89
	18	Trusting–Distrustful	−68	−29	89	00	27	95
10	19	Selective–Undiscriminating	90	85	−35	20	07	89
	20	Tolerant–Choosy	−61	−09	91	−11	−06	85
11	21	Firm–Lax	85	54	−52	34	50	94
	22	Lenient–Severe	−74	−21	88	−12	−33	95
12	23	Discreet–Indiscreet	71	88	03	−30	18	90
	24	Frank–Secretive	−46	−23	61	58	27	83
13	25	Individualistic–Conforming	34	35	−09	80	−10	78
	26	Cooperative–Uncooperative	−36	19	90	−14	07	86
		Ideas and Ability						
14	27	Pragmatic–Unrealistic	63	80	02	11	−06	66
	28	Idealistic–Opportunistic	−19	−06	43	−08	57	52
15	29	Cultivated–Naive	67	81	−06	23	−15	74
	30	Natural–Artificial	−45	−07	74	33	11	67
		Sets of three						
		Temperament						
16	31	Thorough–Careless	89	90	−21	06	24	91
17	32	Moral–Immoral	50	48	03	−26	67	75
18	33	Curious–Uninquisitive	27	51	17	57	−31	72
		Social						
19	34	Forceful–Submissive	45	15	−47	83	15	95
20	35	Peaceful–Belligerent	−24	29	84	−36	10	93
21	36	Polite–Rude	11	55	60	−37	13	82
		Ideas and ability						
22	37	Intelligent–Stupid	49	80	24	37	−15	87
23	38	Foresighted–Short-Sighted	59	92	26	17	−01	95
24	39	Meditative–Unmeditative	68	92	11	−02	05	86
25	40	Witty–Humorless	−40	01	60	56	−41	83
		% Total variance	(34.2)	30.6	28.4	16.4	8.7	84.1
		% Common variance		36.4	33.8	19.5	10.4	100

Table 5 presents such modified indexes between factors from the two analyses.

The very high similarity between the unrotated factors is indicated by an index of .97 in Table 5. Accordingly, they may both be interpreted as the same balanced descriptive dimension of "tightness versus looseness of control." Table 6 provides illustrative sets of traits that had high loadings on both factors.

The systematic use of related traits in the present study shows unambiguously that this dimension is descriptive. In contrast, it would be more like the usual semantic differential factor if results were available only for one of the two columns of scales in Table 6. There would then be a confounding of favorable evaluation with a descriptive characteristic—either tightness (in the left-hand column) or looseness (in the right-hand column). Further, let us anticipate the evidence of the next section that positive and negative traits are typically moderate and extreme. The two columns might then be called "moderate tightness versus extreme looseness" and "moderate looseness versus extreme tightness." A single column, considered separately, would still not be evaluative, although it would then be confounded with evaluation.

It would then also resemble the rotated scale factors. Indeed, the large loadings of the first rotated factor (i) closely resemble the left-hand column of Table 6. The indexes of Table 5 also suggest that this factor is comparable to part of Factors I_U or I, and might be interpreted as "moderate tightness versus extreme looseness," or more specifically as "moderate control versus extreme expression." In a similar way, Factor iii might be interpreted as "moderate assertiveness versus extreme unassertiveness" (corresponding to part of Factor II), and Factor iv as "moderate decidedness versus extreme undecidedness" (corresponding to part of Factor III). There remains Factor ii as the main representative of all the contrasting scales. This factor—which has some resemblance to the right-hand column of Table 6—is most strongly related to the Unrotated Factor I_U and the Rotated Factor III. It might accordingly be interpreted as "moderate looseness versus extreme tightness" or more specifically as "moderate undecidedness versus extreme decidedness."

TABLE 5

COMPARISON BETWEEN THE TWO ANALYSES: MODIFIED INDEXES OF FACTORIAL SIMILARITY

Analysis of scales	Analysis of antecedent terms			
	Unrotated 1st factor I_U	Rotated factors		
		I	II	III
Unrotated 1st factor i_u	.97	.91	.17	.77
Rotated factors				
i	.73	.88	−.11	.42
ii	−.73	−.45	−.47	−.73
iii	.01	.12	.80	−.02
iv	.37	.15	−.16	.71

example, the loading of .72 for the scale *cautious-rash* on the unrotated first factor (i_u). The related loadings from Factor I_U of the other analysis are .76 for CAUTIOUS and −.44 for RASH. The sign of the latter loading was reversed to +.44 since it would otherwise represent the opposite direction from the scale. The two loadings were then averaged: $\frac{.76 + .44}{2} = .60$. This average was compared with the scale loading of .72 in calculating the modified index between the two factors.

TABLE 6

REPRESENTATIVE SCALES WITH LARGE LOADINGS ON THE UNROTATED FIRST FACTOR

Set no.	Scale	Loading	Scale	Loading
2	Self-Controlled–Impulsive	.73	Uninhibited–Inhibited	−.65
3	Serious–Frivolous	.85	Gay–Grim	−.77
8	Thrifty–Extravagant	.89	Generous–Stingy	−.71
9	Skeptical–Gullible	.85	Trusting–Distrustful	−.68
10	Selective–Undiscriminating	.90	Tolerant–Choosy	−.61
11	Firm–Lax	.85	Lenient–Severe	−.74

Note.—This table includes all the sets where both scales have absolute loadings greater than .60 on the unrotated first factor (i_u) in the analysis of scales. In addition, the same sets are the only ones where all four terms, when considered separately, also had absolute loadings greater than .60 on the unrotated first factor (I_U) in the analysis of antecedent terms.

As has just been seen, the evidence is unambiguous that these rotated scale factors are not evaluative: where a scale has a high positive loading, the related scale has a contrasting loading that is low and typically negative. The same conclusion is likely to apply to many other semantic differential factors that have been claimed as evaluative. But the unambiguous evidence is lacking since no method was used to remove the confounding. An example would be the factor analysis of traits by Osgood and Ware, where Osgood (1962) reported finding "such large evaluative factors . . . that little else could be determined [p. 25]." Ware has reported the actual scales involved.[9] For example, the first factor interpreted as evaluative had loadings greater than .70 on the following scales (reordered here): *friendly-hostile, compliant-contrary, soft-hard, sociable-unsociable*; *pleasing-irritating, reputable-disreputable.* This factor is evidently not evaluative but a descriptive dimension confounded with evaluation (particularly as represented by two relatively pure evaluative scales). The descriptive dimension represents interpersonal cooperation versus hostility (it resembles part of Factor ii in the analysis here, where it might have been called "moderate unassertiveness versus extreme assertiveness"). It is easier to interpret factors as evaluative when no method is used to remove the confounding with some descriptive dimension.

Application to Separate Items

The general descriptive dimensions suggested by factor analysis can now be applied back to the analysis of separate items. This analysis considered whether the evaluative similarity of an inference could be accounted for as a by-product of its descriptive similarity. The earlier analysis showed that similarity on relatively specific descriptive attributes could account for inferences within sets. It is now possible to show whether similarity on a more general descriptive dimension can account for the evaluative similarity of inferences between sets. The most convenient illustration uses the dimension of "tightness versus looseness," corresponding to the unrotated first factor of both analyses. Any inference was considered where the antecedent term and the scale each had absolute loadings above .60—most of the relevant traits are those represented in Table 6. For some of these inferences, traits similar on tightness-looseness would also be similar in evaluation. Among such inferences (e.g., between traits on the *same* side of Table 6) evaluative similarity occurred on 100% of 272 items (e.g., for Sets 3 and 11, inferences between "serious" and "firm," "frivolous" and "lax," "gay" and "lenient," "grim" and "severe"). For other inferences, traits similar on tightness-looseness would be dissimilar in evaluation. Among such inferences (e.g., traits on *opposite* sides of Table 6) evaluative dissimilarity occurred on 87% of 277 items (e.g., inferences between "serious" and "severe," "frivolous" and "lenient," "gay" and "lax," "grim" and "firm"), and evaluative similarity on only 13% (e.g., inferences between "self-controlled" and "tolerant"). These results show that either evaluative similarity or evaluative dissimilarity may be predicted from the similarity of descriptive direction. Evaluative similarity tends to occur only when it is the by-product of descriptive similarity.

The evidence of this section goes against the objection considered earlier, that the special cases of near synonyms (i.e., inferences within sets) might only be exceptions to the greater general importance of evaluation. The factors do not represent only near synonyms but broad ranges of traits, and the factors are descriptive not evaluative. Moreover, the same kind of very general descriptive similarity accounts for the evaluation of separate inferences between different sets of traits. The evidence is consistent with the alternative interpretation, that descriptive aspects of judgment are always decisive, and evaluation only seems to be important because it is usually confounded with them.

The next section goes one step further, and considers evidence that evaluation itself may be typically based on a descriptive judgment of the degree of extremeness.

ARISTOTLE, EXTREMENESS, AND EVALUATION

Aristotle, and classic Greek thought generally, held that for personality traits, mod-

[9] Personal communication.

eration was desirable and extremeness undesirable. For ordinary people, the same relationships might not be explicitly stated but still constitute implicit principles of the way they make trait judgments. What evidence could be derived from the present study? The following derivation was examined: consider traits that are basically dissimilar in descriptive direction. If negative traits are more extreme, then among such traits, they should be more dissimilar and represent opposite extremes. They should more definitely exclude each other in trait inferences than would comparable positive traits. Accordingly, evaluative dissimilarity should be greater for inferences from negative traits, and evaluative similarity should be greater for inferences from positive traits.

This derivation was supported wherever evidence could be found in the present study: everyday trait judgments seem to follow Aristotle more than Goldwater, treating extremeness as a vice and moderation as a virtue. Two examples can represent the other findings not reported here. Among the critical inferences within sets, it was reported above that all the item means were in the direction of evaluative dissimilarity. But those from negative antecedent terms were significantly more so (the combined item mean was +1.80) than those from positive antecedent terms (the combined item mean was −1.24). On the other hand, for unselected inferences there was an overall majority of 67% evaluative similarity. The majority for positive antecedent terms (75%) was significantly greater than that from negative antecedent terms (62%). These results are all consistent with the hypothesis that negative traits are more extreme.

These results also represent a tendency to infer towards positive traits. Accordingly, they might be given an alternative (and seemingly simpler) interpretation using the "leniency effect"—the classic finding that trait ratings tend to be favorable. However, like the "halo effect," the leniency effect itself needs to be explained—and might be as a derivation from the hypothesis that negative traits typically represent extremes.

In any case, the results suggested getting more direct evidence that negative traits might represent opposite extremes. A limited supplementary study examined directly whether the negative traits within a set were judged more dissimilar. Forty additional subjects, again Swarthmore undergraduates, ranked the similarity of pairs of traits from each of the 15 sets of four. The pair of negative traits was ranked as less similar than the pair of positive traits in 78% of cases overall. The percentage was greater than 50% of subjects for 14 of the 15 sets (the range was from 48% to 98%), and significantly so for 11 of them (at the .01 level by the sign test). These results suggest that negative traits are typically, but not always, treated as extremes. A descriptive judgment of the degree of extremeness thus may typically be the basis for evaluation. Applying a trait to a person involves the judgment that a person is moderate or extreme on a descriptive dimension. This descriptive judgment is then the basis for evaluating the person as more or less desirable.

The evidence is now completed. It argues strongly against the usual emphasis on evaluation: (*a*) When traits are selected to separate similarity in evaluation from that on some descriptive attribute, descriptive similarity is decisive for all the critical inferences; (*b*) when the inferences are factor analyzed, none of the factors are evaluative; (*c*) the general descriptive dimensions suggested by factor analysis account for the evaluative similarity or dissimilarity of inferences in general; (*d*) evidence suggests that evaluation is typically based on a descriptive judgment of the degree of extremeness.

The next section moves on to make some positive proposals about the nature of trait judgments.

DISCUSSION

Heider (1958, p. 5) discussed two possible advantages in studying the everyday cognitions people have about others. The first is that those cognitions may be a source of evidence about actual personality. It is striking that the factors in the present study of trait cognitions are closely analogous to those found repeatedly for measures of actual people. Indeed, the basic problem of the con-

founding between evaluative and descriptive aspects has its analogue in self-report measures of personality. The analogous confounding is between "social desirability" (with its implication that subjects are responding to evaluative aspects of the items), and the descriptive aspects of the statements. Thus, Block (1965) argued cogently that the typical first factor of the MMPI should be interpreted not as "social desirability" but as a decriptive dimension, which he suggested calling "Ego-Resiliency." Similar dimensions, often called "general adjustment," have appeared in other analyses of personality measures. In contrast is the other main factor of the MMPI, which Block called "Ego-Control." This dimension is directly comparable to the large factors found in the present analyses that concerned very general forms of control. The analysis of trait cognitions is thus related to other evidence about the structure of personality.

Of more central importance here is Heider's second point that everyday "naive" psychology is important in its own right. Thus, that our cognitions of others are ordinarily in terms of traits presents an important problem in our "perception of persons," regardless of whether the most useful scientific formulation of personality is in terms of traits. Historically, many positions in psychology from Thorndike to Skinner have rejected traits as too general for an adequate account of personality. Nevertheless, since people ordinarily do use traits in dealing with others, the cognitive problem remains.

Some general principles can be proposed as to the nature of trait cognitions:

Trait judgments are on an implicit scale relating behaviors and situations. Although trait judgments are readily made, they represent quite complex relationships. Similar behavior is judged differently depending on various situational factors. This point is discussed by Heider (1958, Chapter 4) for the attribution of ability, which he shows to depend not on success or failure as such, but on success or failure in relation to the perceived difficulty of the task. The case is similar for other traits. Whether a statable behavior (e.g., investing $10,000) is judged "bold" or "rash" or not depends on several situational variables, such as the perceived probabilities of various gains and losses, and the proportion the investment represents of the person's resources. Similarly, the telling of a joke that would be judged "frivolous" at a formal conference may not be at a party.

The relationships appear to be lawful. Behaviors are treated as having systematic order—for example, as varying in riskiness or seriousness. Situations are treated in a compable way; for example, party, formal conference, funeral. Behavior and situation are then typically related according to principles of contrast—for example, a person is judged "frivolous" for less serious behavior in more serious situations. The result is an overall scale of behavior relative to situation. This scale represents the descriptive directions, and provides a context for a particular judgment.

An alternative approach (implied by Heider, 1958) might bypass this analysis of behaviors relative to situations and simply consider an implicit distribution of persons. The high extreme on a scale of "riskiness," for example, would then not be defined as a function of amount risked relative to total resources and probabilities of success, etc., but simply as a risk that would be taken by only a small proportion of people.

Trait judgments are decisions relative to some criterion on an implicit scale. The extremes are evaluated negatively. A trait judgment is a categorical decision relating the person judged to the implicit scale. The scale is treated as being divided according to some criterion. A particular judgment is that the person is above or below the criterion. A typical scale is treated as having both upper and lower criterion boundaries, that separate "high" and "low" categories from the rest of the scale. These are the extreme categories that are evaluated negatively.

This proposal thus treats trait judgment as analogous to recent decision theories (e.g., of signal detection in psychophysics.) The particular model proposed is represented in Figure 1, and illustrated for traits involving risk taking.

Figure 1 indicates how a trait may represent an answer to two different questions: (a) Is the person above or below the upper

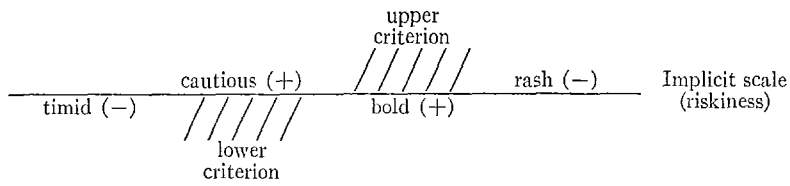

FIG. 1. Proposed model for trait judgments (illustrated for risk-taking).

criterion (e.g., "rash" or "cautious")? (*b*) Is the person above or below the lower criterion (e.g., "bold" or "timid")? Figure 1 presents each criterion as a cross-hatched region, providing for judgments not only of above and below the criterion, but also of uncertainty either way. An alternative would treat each of the four traits as representing a separate and more precise decision between the trait and its exact complement (e.g., "cautious" versus "incautious"; "rash" versus "unrash"). This alternative would have greater generality, but would not permit as compact a representation.

The proposed model provides a straightforward account of many of the main findings: (*a*) In the relation between traits, similarity of descriptive direction is more important than evaluative similarity; (*b*) negative traits represent extreme forms of positive traits. As opposing extremes, they would be more dissimilar and exclude each other more definitely in an inference.

Further implications of the model might be examined by considering possible combinations of traits. Thus, the model indicates that in contrast to negative traits, opposing positive traits (e.g., "cautious" and "bold") might apply to the same instance (e.g., "discretion is the better part of valor").

Finally, the proposed model suggests a way of systematically analyzing the disagreement and agreement between different judges. Psychology has tended to emphasize the disagreement and subjectivity of judgment, and neglect the possible underlying agreement. Actually, different judges often seem to agree about descriptive direction and disagree about evaluation. For example, one well-known investigation (Sargent, 1939) studied the use of parallel terms like "radical" and "progressive" by the *Chicago Tribune* and the *New York Times*. Other examples are suggested by Vinacke's (1949) investigation of the views of different ethnic groups in Hawaii; for example, the Japanese are described by themselves and others as having "close-knit family," but by others as also "clannish"; the haoles (Caucasians) are described generally as "confident" and "outspoken," but by others as also "boastful," "snobbish," and "conceited."

The model implies that a specific judgment might be analyzed into two parts: (*a*) a specific descriptive judgment—that is, locating the object of judgment on the relevant implicit scale; (*b*) a general rule relating descriptive judgment to evaluation—that is, the location of the criterion in the implicit scale. Judges may agree on the specific descriptive judgment and differ about the general criterion. The criteria would be likely to vary like adaptation levels according to the judge's typical experience. His own typical behavior is likely to be on the moderate side of his criteria. For another judge with different criteria, the same behavior might be judged extreme and evaluated negatively. There results what Brown (1965, Chapter 4) has called an "ethnocentrism of evaluation." Those who are relatively "tight" and "loose" may be agreed about this fact, but the "tight" person regards tighter behavior as normal (e.g., "thrifty," "serious") and looser behavior as extreme (e.g., "extravagant," "frivolous"). The "loose" person has different criteria, regarding tighter behavior as extreme (e.g., "stingy," "grim") and looser behavior as normal (e.g., "generous," "gay"). In this way, disagreement about evaluation may be lawfully related to agreement about descriptive direction.

This final section has proposed a theoretical model for judgments using traits. This model represents many of the earlier findings in a lawful relationship between evaluative and

descriptive aspects of judgment: evaluation is a secondary aspect of judgment and is typically based on the extremeness of descriptive judgment.

REFERENCES

ALLPORT, G. W., & ODBERT, H. S. Trait names: A psycholexical study. *Psychological Monographs,* 1936, 47(1, Whole No. 211).

ANDERSON, N. H. Application of an additive model to impression formation. *Science,* 1962, 38, 817–818.

ASCH, S. E. Forming impressions of personality. *Journal of Abnormal and Social Psychology,* 1946, 41, 258–290.

BLOCK, J. *The challenge of response set.* New York: Appleton-Century-Crofts, 1965.

BROWN, R. W. *Social Psychology.* New York: Free Press, 1965.

BRUNER, J. S., SHAPIRO, D., & TAGIURI, R. The meaning of traits in isolation and in combination. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford: Stanford University Press, 1958. Pp. 277–288.

CATTELL, R. B. *The description and measurement of personality.* Yonkers, N. Y.: World Book Company, 1946.

CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin,* 1955, 52, 177–193.

FISHBEIN, M., & HUNTER, R. Summation versus balance in attitude organization and change. *Journal of Abnormal and Social Psychology,* 1964, 69, 505–510.

HARMAN, H. H. *Modern factor analysis.* Chicago: University of Chicago Press, 1960.

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

LAFORGE, R., & SUCZEK, R. The interpersonal dimension of personality: III. An interpersonal checklist. *Journal of Personality,* 1955, 24, 94–112.

LEARY, T. F. *Interpersonal diagnosis of personality.* New York: Ronald Press, 1957.

OSGOOD, C. E. Studies on the generality of affective meaning systems. *American Psychologist,* 1962, 17, 10–28.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: University of Illinois Press, 1957.

OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in the prediction of attitude change. *Psychological Review,* 1955, 62, 42–55.

PODELL, J. E. A comparison of generalization and adaptation-level as theories of connotation. *Journal of Abnormal and Social Psychology,* 1961, 62, 594–597.

PODELL, J. E., & AMSTER, H. Evaluative concept of a person as a function of the number of stimulus traits. *Journal of Personality and Social Psychology,* 1966, 4, 333–336.

ROSENBERG, M. J., & ABELSON, R. P. An analysis of cognitive balancing. In M. J. Rosenberg et al. (Eds.), *Attitude organization and change.* New Haven: Yale University Press, 1940. Pp. 112–163.

SARGENT, S. S. Stereotypes and the newspapers. *Sociometry,* 1939, 2, 69–75.

STEINER, I. D. Ethnocentrism and tolerance of trait inconsistency. *Journal of Abnormal and Social Psychology,* 1954, 49, 349–354.

VINACKE, W. E. Stereotyping among national-racial groups in Hawaii: A study in ethnocentrism. *Journal of Social Psychology,* 1949, 30, 265–291.

(Received September 9, 1966)