The Prediction of Semantic Consistency in Self-Descriptions: Characteristics of Persons and of Terms That Affect the Consistency of Responses to Synonym and Antonym Pairs

Lewis R. Goldberg
University of Oregon

James M. Kilkowski
Los Angeles, California

Subjects described themselves, using an alphabetically ordered list of 191 trait adjectives, which included sets of synonyms and antonyms, half of each type more difficult than the other half. Subjects were randomly assigned to one of two experimental conditions. In one condition, each adjective was listed with its dictionary definition; in the other condition, only the adjectives were listed. All subjects were administered a battery of demographic, cognitive, and personality measures. We analyzed both the relative consistency elicited by different pairs of terms and the individual differences in semantic consistency displayed by different sorts of subjects. Although the provision of definitions served to increase consistency (especially for the difficult antonyms), it did not decrease the range of consistency values across either synonym or antonym pairs. And, although interpair differences in semantic consistency were as difficult to predict in this study as in previous ones, individual differences were highly predictable. The implications of our many findings are discussed in the context of various hypotheses about semantic inconsistency in self-reports.

Over the years, an important theoretical controversy has centered on the extent to which the correlational structures discovered in self- and peer-rating studies (e.g., Cattell, 1957; Norman, 1963) reflect primarily (a) the organization of personality traits among the ratees or (b) the raters' shared "implicit personality theory" (see Norman & Goldberg, 1966). Indeed, some investigators have proposed that the results of such studies may have *no* bearing on actual personality trait covariation, but merely reflect the way in which the raters generally view the interrelations among such traits (D'Andrade, 1965, 1974; Levy & Dugan, 1960; Mulaik, 1964; Shweder, 1975, 1977, 1979, 1982; Shweder & D'Andrade, 1980).

It is likely that this issue can never be resolved in the absence of detailed information about the sheer similarity of meaning among the trait-descriptive terms used in these investigations. To determine the extent to which dictionary-defined similarity of meaning is mirrored in the actual use of trait terms when people describe themselves or others, we have been studying the most simple cases (Goldberg, 1978a): (a) *synonyms* (or *near synonyms,* if one assumes that no two orthographically different forms are ever exact

, Support for this study was provided by Grant MH-12972 from the National Institute of Mental Health, U.S. Public Health Service. Some of the hypotheses were originally formulated by our colleagues, including Mark Layman, Warren T. Norman, Leonard G. Rorer, and Jerry S. Wiggins. This report was written while the first author was a Fellow-in-Residence at the Netherlands Institute for Advanced Study (NIAS), supported by funds from NIAS and from the James McKeen Cattell Fund.

Preliminary analyses of these data were carried out by the second author and reported in his doctoral dissertation at the University of Minnesota. The analyses described in this report were carried out by the first author, with the assistance of Frank B. Brokken of the University of Groningen in The Netherlands. The report has benefitted substantially from the thoughtful critiques of earlier versions provided by Matthias Burisch, David M. Buss, William F. Chaplin, Roy D'Andrade, Sarah E. Hampson, Oliver P. John, Adrienne Lehrer, Clarence McCormick, and Richard A. Shweder.

Requests for reprints should be sent to Lewis R. Goldberg, Institute for the Measurement of Personality, 1201 Oak Street, Eugene, Oregon 97401.

Journal of Personality and Social Psychology, 1985, Vol. 48, No. 1, 82–98

equivalents in all contexts), as listed in dictionaries and synonym finders; and (b) *antonyms* (including root-negation pairs), as listed in some dictionaries. For pairs of synonyms and of antonyms, we have examined two major kinds of relations inferred from self-descriptions: (a) *individual differences in consistency* (e.g., a correlation for each subject across the pairs of terms) and (b) *intertrait differences in consistency* (e.g., a correlation between the two terms in each pair across the sample of subjects). Each kind of index has been used in an iterative fashion to help cull both the subject sample and the set of synonym or antonym pairs. Subjects who respond in a grossly inconsistent fashion are probably taking the task less seriously than those whose consistency is more modal. Analogously, synonym or antonym pairs that elicit grossly inconsistent responses from the typical subject may have initially been inappropriately paired. Nonetheless, after the exclusion of careless subjects and the culling of any erroneously paired terms, there typically remains a considerable range of individual differences in consistency in the sample of subjects, and a substantial range of intertrait differences within the set of synonym or antonym pairs.

Individual differences in semantic consistency are somewhat general across different sets of synonyms and antonyms embedded in the same questionnaire. For example, using the 1,710 trait adjectives included in one of our inventories (Goldberg, 1982), we selected 339 synonym pairs and 110 antonym pairs from Webster's *Dictionary of Synonyms*. For each subject in one sample, a correlation across the synonyms and another across the antonyms were derived from the subject's self-ratings. The correlation between these two indexes, across the 187 subjects, was .68 ($p < .001$); moreover, the substantial congruence between these two consistency indexes was manifested despite the fact that the sample had already been culled of all grossly inconsistent responders. For such refined samples, individual differences in synonym consistency correlations typically range from .10 to .70, with a mean between .45 and .50; individual differences in antonym consistency generally cover much the same range, with slightly lower means (about −.40).

In contrast, intertrait correlations tend to cover a wider range, and their means are typically closer to zero; some examples are as follows: (a) Among the 110 antonym pairs, the consistency correlations across the 187 subjects ranged from .10 (*ingenuous–cunning* and *icy–fiery*) to −.90 (*masculine–feminine*), with a mean of −.19; (b) among the 399 synonym pairs, the consistency correlations ranged from −.05 (*objective–impartial* and *slow–leisurely*) to over .90 (*feminine–womanly* and *masculine–manly*), with a mean of .33; (c) among a set of 389 root-negation pairs, the consistency correlations covered the range from −.10 (*selfless–selfish*) to over −.85 (*feminine–unfeminine* and *masculine–unmasculine*), with a mean of −.48.

Over the years, we have used two approaches, "clinical" and "statistical," to predict intertrait differences in semantic consistency. From time to time, we have held contests and offered prizes to see how accurately subjects could predict the rank order of a set of 15 to 20 representative synonym or antonym pairs. Few of the highly motivated subjects in these various contests have been able to predict the actual ranking of intertrait differences at a statistically significant level. Overall, subjects' judgments correlate near zero with the true ranks.

Statistical analyses, on the other hand, have been more helpful. Such analyses have shown that two predictors—the squared difference between the terms' mean ratings and the midpoint of the rating scale, and the squared difference between the mean ratings of the two terms—typically are predictive of intertrait consistency with multiple correlations in the .50 to .70 range. These findings indicate that (a) the more extreme the mean values of the terms are, the higher (more consistent) are the correlations; (b) for synonyms, the closer the mean ratings of the two terms are to each other, the higher (more consistent) are the correlations; and (c) for antonyms, the farther apart the mean ratings of the two terms are, the higher (more consistent) are the correlations. However, even when these effects are controlled, it is possible to assemble sets of synonyms and antonyms that cover a substantial range of consistency values. Table 1 provides an example of 15 such antonyms.

Table 1
Representative Set of Antonyms Illustrating the Range of Intertrait Differences in Consistency Correlations That Have Been Found

	M		*SD*			*M*		*SD*		Correlation between Terms A and B	
Term A	U.S.	Aus	U.S.	Aus	Term B	U.S.	Aus	U.S.	Aus	U.S.	Aus
Silent	4.3	4.2	2.0	2.0	Talkative	5.8	5.9	2.0	1.8	−.55	−.49
Optimistic	6.6	6.4	1.7	1.6	Pessimistic	4.1	4.3	2.2	2.0	−.45	−.63
Decisive	5.8	6.1	1.7	1.8	Irresolute	4.1	4.2	1.7	1.7	−.38	−.37
Cheerful	7.0	6.8	1.2	1.3	Gloomy	3.3	3.3	1.5	1.6	−.36	−.30
Ignorant	2.5	3.0	1.4	1.4	Informed	6.6	6.4	1.3	1.2	−.30	−.28
Frugal	4.9	4.4	1.9	1.8	Wasteful	3.6	4.2	1.8	1.8	−.22	−.29
Acquiescent	5.0	4.9	1.6	2.0	Rebellious	4.9	4.8	2.0	1.8	−.20	−.20
Despondent	3.6	4.0	1.6	1.7	Light-hearted	6.0	6.1	1.6	1.4	−.20	−.15
Lewd	2.9	2.9	1.7	1.9	Chaste	4.4	4.9	2.2	2.4	−.18	−.16
Impulsive	5.8	5.4	1.8	1.8	Deliberate	6.0	6.3	1.5	1.4	−.17	−.14
Affirmative	6.5	5.9	1.2	1.7	Negative	3.6	3.6	1.7	1.6	−.12	−.16
Sedate	4.3	4.6	1.9	1.8	Flighty	4.0	3.6	1.8	1.7	−.09	−.09
Lenient	6.5	6.1	1.4	1.4	Stern	4.2	4.1	1.8	1.6	−.08	−.02
Foolhardy	3.5	3.7	1.8	1.7	Wary	5.3	6.0	1.8	1.4	−.02	.03
Obstinate	4.9	5.2	1.9	1.9	Pliable	5.7	5.5	1.8	1.5	.03	.03

Note. These values are based on the responses to the 1,710 self-description inventory from a sample of 187 University of Oregon undergraduates (U.S.) and from a sample of 123 undergraduates at the University of Western Australia (Aus). (From Goldberg, 1978a).

Before attempting to explain these intertrait differences in semantic consistency, we must determine the extent to which they generalize across diverse samples of subjects. To find out, we have compared both synonym and antonym correlations in three samples: (a) 100 University of Michigan undergraduates (Norman, 1967), (b) 187 University of Oregon undergraduates, and (c) 123 University of Western Australia undergraduates. Although there were occasional trait pairs that differed substantially between two samples (the most extreme example is the antonym pair, *bashful–forward,* which elicited consistency correlations of −.43 in the Oregon sample and .04 in the Australian sample, despite virtually identical means and standard deviations of the ratings of the two terms in both samples), intersample congruence correlations were all in the .80 to .90 range.

There are a number of possible explanations for the large intersubject and intertrait differences in semantic consistency, two of which we tested in the present study. The first, *response biases,* focuses primarily on characteristics of the subjects, and the second, *word difficulty,* on characteristics of the task itself.

Response Biases

The effects of individual differences in social desirability response bias and of "acceptance acquiescence" (Bentler, Jackson, & Messick, 1971) on the responses to synonym and antonym pairs are not identical. The social desirability values of the two terms in a synonym pair are quite similar, whereas the two terms in an antonym pair have opposite desirability values. Consequently, individual differences in social desirability will tend to increase both the positive correlations between synonyms and the negative correlations between antonyms; such individual differences should not produce differential effects when synonym pairs are compared with antonym pairs. On the other hand, the effects of acquiescence would produce higher correlations among all pairs of terms. Such a bias would raise the positive correlations between synonyms and would attenuate the negative correlations between antonyms. Thus the invocation of these two response biases helps to explain the findings that (a) in self-descriptions the absolute magnitudes of the correlations between synonym pairs are higher than those between antonym pairs, and (b)

nonsynonymous terms of similar evaluation tend to correlate more highly than one might expect on the basis of their semantic content alone.

Word Difficulty

Academicians in general and psychologists in particular may fool themselves by assuming that the typical college student has as high a level of verbal facility as they had, or think they had, when they were undergraduates. Perhaps the major difference found between trait pairs in synonym and antonym consistency reflects the differential popularity of those terms in contemporary discourse. Moreover, the difficulty indexes provided by Norman (1967), based on University of Michigan students, may vastly underestimate the difficulty level of those terms among samples of subjects from less selective schools. One implication of this hypothesis is that if one provided subjects with the dictionary definitions of each term, the semantic consistency of their self-descriptions should increase significantly.

The present study provides a direct test of these hypotheses. Specifically, we made four predictions: (a) Providing dictionary definitions of terms should significantly increase synonym and antonym consistency over a condition in which they are not provided. (b) Subjects should respond more consistently to common terms than to those of greater difficulty; however, this factor should be of significance only when definitions are not provided, thus resulting in an interaction between the experimental conditions and difficulty level. (c) Subjects of higher verbal intelligence should respond more consistently than those of lower intelligence; again this factor should matter more when no definitions are available, thus leading to a significant interaction between experimental condition and intelligence level. (d) Consistency should be greater for synonym than for antonym pairs, but these differences should be reduced (if not eliminated) when the effect of acquiescence response bias is controlled.

Method

Overview

Subjects described themselves, using an alphabetically ordered list of 191 trait adjectives. Included in this list were sets of synonyms and of antonyms, half of each type more difficult than the other half. Subjects were randomly assigned to one of two experimental conditions. In one condition, each adjective was listed with its dictionary definition; in the other condition, only the adjectives were listed. Subjects in both conditions were administered a battery of demographic, cognitive, and personality measures. Analyses were focused both on the prediction of the relative consistency elicited by different pairs of terms and on individual differences in semantic consistency displayed by different sorts of subjects.

The Stimulus Set

Within the subset of 1,710 trait-descriptive adjectives (Goldberg, 1982), which had been culled mainly from the 2,800 trait terms analyzed by Norman (1967), 339 synonym pairs and 110 antonym pairs are listed in Webster's *Dictionary of Synonyms*. We graded each of these pairs for difficulty level, using the data provided by Norman (1967) from 50 male and 50 female college students. Specifically, the number of the 100 subjects who reported that they did not understand the meaning of the term was summed across the two terms in each synonym and antonym pair. All the pairs were rank ordered on this composite index of difficulty, and then split at the median.

The cutoff criterion between the two subsets of pairs turned out to be quite low by absolute standards. If the summed difficulty-level index was three or higher, the pair was designated as "difficult"; otherwise it was "easy." In other words, if 3% or more of the 100 subjects indicated that they did not know the meaning of one term in a pair (and none so indicated for the other term), then that pair fell into the "difficult" subset. Such a low cutoff criterion can be rationalized on the grounds that subjects generally tend to overestimate their knowledge of the meaning of trait adjectives. On the other hand, it is important to keep in mind that the experimental distinction used in this study between difficult and easy pairs is only a relative one, and that on the average the difficult pairs were still quite familiar when compared with all trait-descriptive terms in the English language (e.g., Allport & Odbert, 1936).

From each of the four sets of pairs (Synonyms vs. Antonyms × Difficult vs. Easy Pairs), 20 pairs were randomly selected, subject to two restrictions: (a) Terms referring either to religion (e.g., *reverent–irreverent*) or to gender (e.g., *manly–virile*) were omitted, because previous research has demonstrated that semantic consistency to such pairs is extremely high in samples composed of both sexes and of persons of heterogeneous religiosity; and (b) all pairs derived from the same root (e.g., *responsible–irresponsible, timid–timorous*) were excluded so as to ensure that semantic consistency would not be confounded with similarity based on a common root or with word order in an alphabetically arranged list of terms.

The final set of terms included 20 pairs (40 terms) each of easy synonyms, difficult synonyms, easy antonyms, difficult antonyms, all of which are listed in Table 2. These 160 critical terms were embedded alphabetically in a total list of 191 trait adjectives. Definitions for each of the 191 terms were selected from *Webster's Third*

Table 2
Synonym and Antonym Pairs, Their Difficulty Level, and Their Semantic Consistency in Each Experimental Condition

Synonyms	Diff A	Diff B	Consist DEF	Consist NON	Antonyms	Diff A	Diff B	Consist DEF	Consist NON
				Easy pairs					
Economical–thrifty	0	1	77	69	Optimistic–pessimistic	0	0	−61	−64
Responsive–sympathetic	0	0	66	21	Silent–talkative	0	0	−59	−53
Benevolent–charitable	0	0	66	20	Conscientious–unscrupulous	1	1	−55	−16
Brave–courageous	0	0	58	73	Bold–cowardly	0	0	−53	−19
Critical–faultfinding	0	0	58	55	Deliberate–impulsive	0	0	−48	−10
Rigorous–strict	1	1	58	34	Brisk–sluggish	0	0	−45	−12
Moderate–temperate	0	1	57	38	Easygoing–irritable	0	1	−45	−25
Conventional–formal	2	0	57	34	Profound–shallow	2	0	−42	−31
Fiery–spirited	0	0	57	30	Calm–stormy	0	1	−39	−27
Faithful–loyal	0	0	55	72	Timid–venturesome	0	0	−38	−22
Affectionate–loving	0	1	54	72	Awkward–graceful	0	0	−37	−40
Gentle–mild	0	1	49	39	Assertive–retiring	1	1	−35	−15
Obstinate–stubborn	1	1	46	41	Deceitful–trustworthy	2	0	−29	−34
Impartial–unbiased	0	0	45	53	Blunt–tactful	0	0	−29	−21
Frolicsome–playful	1	0	41	45	Constant–fickle	2	0	−29	−34
Submissive–tame	1	0	37	27	Indirect–straightforward	1	0	−23	−48
Accurate–precise	0	0	35	41	Fanciful–realistic	1	0	−17	04
Sophisticated–worldly-wise	0	1	23	22	Humble–proud	0	0	−05	−32
Daring–reckless	0	0	03	33	Exacting–lenient	0	2	−01	15
Chivalrous–gallant	1	0	−06	38	Abusive–complimentary	1	0	01	−23
				Difficult pairs					
Extravagant–lavish	1	4	77	53	Obtrusive–shy	17	0	−62	−08
Revengeful–vindictive	0	12	72	57	Industrious–slothful	0	9	−50	−21
Testy–touchy	14	0	66	32	Decisive–irresolute	0	10	−50	−33
Sly–wily	0	14	61	42	Affable–reserved	24	0	−49	12
Arrogant–haughty	1	4	61	28	Frugal–wasteful	6	0	−46	−19
Judicious–prudent	2	3	59	46	Energetic–lethargic	1	22	−45	−39
Pensive–reflective	7	1	58	36	Languid–vivacious	23	2	−44	−02
Miserly–niggardly	0	18	53	43	Cooperative–factious	0	18	−39	−01
Fastidious–finicky	16	3	51	13	Cheerful–doleful	0	15	−37	−31
Headlong–impetuous	5	10	49	43	Callous–tender	4	0	−37	−12
Gregarious–social	19	0	48	36	Amicable–antagonistic	16	1	−37	11
Closefisted–stingy	8	0	45	42	Earnest–frivolous	0	6	−36	−13
Credulous–gullible	10	0	45	15	Flighty–sedate	0	3	−31	14
Apathetic–impassive	3	4	42	14	Audacious–circumspect	7	25	−30	01
Brazen–impudent	2	13	37	29	Amiable–surly	1	22	−28	−16
Buoyant–effervescent	2	7	35	40	Capricious–steadfast	18	4	−21	03
Brusque–curt	18	8	33	45	Cunning–ingenuous	0	11	−19	06
Cautious–wary	0	4	24	06	Acquiescent–rebellious	18	0	−12	−03
Forbearing–tolerant	10	0	16	24	Docile–unruly	7	0	−12	02
Imperturbable–nonchalant	7	1	14	27	Coy–pert	0	9	−06	14

Note. Diff = difficulty values for both terms in each pair, which are percentages of subjects who did not know the meaning of the term (Norman, 1967); Consist = semantic consistency values, based on the continuous (1–9) response scale and the correlation coefficient (decimals omitted) as the index of consistency; A = first word in pair, B = second word in pair; DEF = definition condition; NON = nondefinition condition.

Unabridged Dictionary; when more than one definition was given under an entry, the one most appropriate for personality description was selected. Care was taken to avoid inducing an "index correlation" between the two terms in a synonym pair (and a reverse index correlation between the two terms in an antonym pair) resulting from the use of the same expression in the definitions of the two terms in any pair.

Subjects in both conditions rated the accuracy of each term as a self-descriptor, using the following rating scale: 1 = *extremely inaccurate,* 2 = *very inaccurate,* 3 = *moderately inaccurate,* 4 = *slightly inaccurate,* 6 = *slightly accurate,* 7 = *moderately accurate,* 8 = *very accurate,* and 9 = *extremely accurate;* no middle (5) response option was provided. This rating format permitted analyses of semantic consistency based on the continuous (1–9) rating scale, as well as on a dichotomized scale transformation (1–4 vs. 6–9).

Subjects

Subjects, recruited through a newspaper advertisement, were all well-paid volunteers. Of the 160 subjects so recruited, 80 were randomly assigned to the definition condition and 80 to the nondefinition condition. The total sample of subjects included 95 women and 65 men; their ages ranged from 15 (one subject) to 73 (one subject), with a mean of 27. Years of education ranged from tenth grade to four years postcollege, the average subject having completed two years of college.

Measures of Individual Differences

Each subject completed (a) a short demographic questionnaire that included variables such as sex, age, and educational level; (b) the first section (synonyms and antonyms) of Terman's (1956) Concept Mastery Test (CMT); (c) Snyder's (1974) Self-Monitoring Scale; and (d) the 480-item California Psychological Inventory (CPI). These measures were selected so as to include (a) the most obvious sort of demographic controls, (b) a measure of verbal fluency (CMT), (c) a measure of self-presentation strategies (Self-Monitoring Scale), and (d) a broad bandwidth inventory of personality characteristics (CPI).

From the CPI item pool, 20 scales were scored: Dominance (Do), Capacity for Status (Cs), Sociability (Sy), Social Presence (Sp), Self-Acceptance (Sa), Well-Being (Wb), Responsibility (Re), Socialization (So), Self-Control (Sc), Tolerance (To), Good Impression (Gi), Communality (Cm), Achievement via Conformance (Ac), Achievment via Independence (Ai), Intellectual Efficiency (Ie), Psychological-Mindedness (Py), Flexibility (Fx), Femininity (Fe), Empathy (EMP; Hogan, 1969), and Autonomy (AUT; Kurtines, 1978).

Indexes of Semantic Consistency

Because consistency to pairs of synonyms and antonyms can be measured in diverse ways, it is important to ascertain the extent to which findings based on any one such index generalize to others. In the present study, indexes based on the continuous (1–9) response format were compared with those based on the dichotomized (e.g., true–false) format. Within each format we used two types of similarity indexes, correlation coefficients and distance measures. Thus we could compare the results from four different measures of semantic consistency: (a) correlations across all of the values of the continuous rating scale, (b) phi coefficients (the correlations based on the dichotomized response scale), (c) d^2 (the squared difference in response values to the two terms in each synonym pair, and to the two terms in each antonym pair after the response values for one term of the pair have been reflected), and (d) the proportion of semantically consistent responses based on the dichotomized response scale. (For this last index, responses are consistent if both terms of a synonym pair were, or both were not, rated as self-descriptive; for antonym pairs, responses are consistent if one term was rated as self-descriptive and the other was not.)

We focus our findings primarily on those two of the four consistency indexes that are most likely to be used in other studies: *r*, the correlation coefficient, which was based on the continuous response format, and %, the consistency proportions, which were based on the dichotomized response format. (Results based on the other two indexes are available from the first author.)

Analyses

Two major types of analyses were carried out: (a) analyses of interpair differences in consistency within each of the two synonym and each of the two antonym sets, and (b) analyses of individual differences in semantic consistency within each of the two samples of subjects. For these latter analyses, values for the correlational index of consistency are always r_c coefficients (Cohen, 1969), the Pearson correlation across the $2 \times k$ pairs (k equals the number of pairs) constructed by including each pair twice, once in each term order (e.g., *daring–reckless* and *reckless–daring*).

Results

Analyses of Differences Among the Synonym and Antonym Pairs

What is the effect of providing definitions on the semantic consistency elicited by the average pair in each of the four sets? In Table 3 we provide one answer to this question that is based on the correlation as an index of consistency. Included in this table are both the mean and median consistency values, plus the ranges, across the 20 pairs of each type, in the definition and the nondefinition conditions, as well as the difference between the mean values in the two conditions. The results confirmed our hypotheses: (a) Consistency was greater for synonym than for antonym pairs in both conditions. Specifically, in the definition condition, the mean consistency correlations (Column 1 of Table 3) were .47 for both easy and difficult synonyms,

as compared with mean values of −.34 and −.35, respectively, for the two types of antonyms; in the nondefinition condition, the mean consistency correlations were .43 and .34 for the synonyms, as compared to −.25 and −.07 for the antonyms. (b) Also as hypothesized, the effect of providing definitions increased semantic consistency least for the easy synonyms (.04) and most for the difficult antonyms (.28). Nonetheless, the most surprising aspect of these findings is the considerable *range* of consistency values elicited within the 20 pairs of each type, even in the definition condition. In fact, for the easy synonyms the provision of definitions served to increase the range of semantic consistency; in this condition, the consistency correlations varied from .77 (*economical–thrifty*) to −.06 (*chivalrous–gallant*).

Indeed, if one were to have classified the pairs by difficulty level solely on the basis of the consistency they elicited in the present study, the relation between the a priori classification and the empirical one would not be high. Specifically, on the basis of the correlational index of consistency, the concordance percentages between the two classifications were only .55 for the 40 synonym pairs and .75 for the 40 antonym pairs. On the basis of the dichotomous index, the concordance percentages were only .55 for both the synonyms and the antonyms. (Concordance percentages of .50 would be expected by chance.)

In the upper left quadrant of Table 4 the mean values displayed in the first column of Table 3 are reproduced, so as to permit their comparison with other kinds of consistency indexes. In order to discover the effect of individual differences in response acquiescence on the consistency correlations, we present in the upper right quadrant of Table 4 the average correlation after the response protocols of each subject were first standard (Z) scored, thus ensuring that all subjects had the same mean and variance of their response distribution across the terms. A comparison of the upper left and right quadrants of the table permits a direct specification of the effects on semantic consistency of

Table 3
Effect of Providing Definitions on the Semantic Consistency of the Average Synonym and Antonym Pair and of the Average Subject in Each Experimental Condition

Pair type and experimental condition	Mean consistency		Median consistency		Range of values	
	Pairs	Individuals	Pairs	Individuals	Pairs	Individuals
Easy synonyms						
DEF	.47	.52	.54	.55	.77 to −.06	.87 to −.04
NON	.43	.41	.38	.41	.73 to .20	.81 to −.43
Difference	.04	.11	.16	.14		
Difficult synonyms						
DEF	.47	.61	.48	.65	.77 to .14	.94 to −.05
NON	.34	.33	.36	.33	.57 to .06	.86 to −.41
Difference	.13	.28	.12	.32		
Easy antonyms						
DEF	−.34	−.49	−.38	−.55	−.61 to .01	−.83 to .23
NON	−.25	−.39	−.24	−.45	−.64 to .15	−.78 to .18
Difference[a]	.09	.10	.14	.10		
Difficult antonyms						
DEF	−.35	−.56	−.37	−.61	−.62 to −.06	−.89 to .18
NON	−.07	−.24	−.02	−.23	−.39 to .14	−.68 to .33
Difference[a]	.28	.32	.35	.38		

Note. DEF = definition condition; NON = nondefinition condition. All values are based on the continuous response scale and the correlation coefficient as an index of consistency. The synonym and antonym correlations were computed across the 80 subjects in each condition. The individual differences correlations are r_c coefficients (Cohen, 1969) computed across the 20 × 2 = 40 pairs of a type.

[a] Absolute values.

Table 4
Effect of Providing Definitions on Semantic Consistency: Comparison Among Different Consistency Indexes

Type	Condition			Condition		
	DEF	NON	D − N	DEF	NON	D − N
Continuous Response Scale: Correlation (*r*) index	Original responses			*Z*-scored responses		
Easy synonyms	.47	.43	.04	.45	.37	.08
Difficult synonyms	.47	.34	.13	.46	.27	.19
Easy antonyms	−.34	−.25	.09	−.38	−.29	.08
Difficult antonyms	−.35	−.07	.28	−.40	−.15	.25
Dichotomized Response Scale	Phi coefficient			% Consistent		
Easy synonyms	.36	.32	.04	.75	.73	.02
Difficult synonyms	.42	.26	.15	.78	.66	.12
Easy antonyms	−.22	−.14	.08	.67	.60	.07
Difficult antonyms	−.25	−.04	.20	.70	.57	.13

Note. All values are based on the average synonym or antonym pair, when computed across the 80 subjects in the Definition (DEF) or the 80 subjects in the Nondefinition (NON) conditions. D − N = difference; these are absolute values.

individual differences in acquiescence. Note that by our removing acquiescence differences by *Z* scoring, the consistency correlations decreased slightly for synonyms and increased slightly for antonyms, precisely the effects that would be predicted from an acquiescence hypothesis. However, both effects were small, the differences in correlations ranging from .01 to .07 for the synonyms and from .04 to .08 for the antonyms. Because synonym consistency remained higher than antonym consistency even after *Z* scoring, response acquiescence can not be invoked as the sole determinant of the differences in semantic consistency between synonyms and antonyms.

In the lower two quadrants of Table 4 we present the consistency values for two indexes (phi and %) based on the dichotomized response scale. The values in the lower left quadrant (phi coefficients) can be compared directly with those in the upper left quadrant, because both are based on the correlational index of consistency; not surprisingly, all of the phi coefficients are smaller than the analogous correlations that are based on the continuous response scale. Of greater importance, the pattern of the eight consistency values displayed in all four quadrants of Table 4 are virtually identical, demonstrating that this pattern of results is not an artifact of using a particular measure of semantic consistency.

Although the semantic consistency elicited by the average synonym and antonym pair was increased by the provision of definitions, this general effect did not hold for every pair. In Table 5 we list those pairs that were affected most dramatically by the provision of definitions. In the left half of the table are examples of pairs that elicited far greater consistency when definitions were provided than when they were not; in the right half are listed some pairs that actually elicited less consistency in the definition than in the nondefinition conditions.

Can these enigmatic findings be predicted from other response characteristics of the two terms in each pair? In Table 6 we present the correlations between each of five possible predictors of semantic consistency and the actual consistency values; correlations are presented for both experimental conditions, and separately for the synonym and the antonym pairs. The first two of these five potential predictors have been shown in past studies to show some relation with semantic consistency: (a) The smaller the difference in means between the two terms in a synonym

Table 5
Examples of Synonym and Antonym Pairs for Which Consistency Changed the Most When Definitions Were Provided

Consistency increased				Consistency decreased			
Pairs	DEF	NON	D − N	Pairs	DEF	NON	D − N
			Easy synonyms				
Benevolent–charitable	.66	.20	.46	Chivalrous–gallant	−.06	.38	−.44
Responsive–sympathetic	.66	.21	.45	Daring–reckless	.03	.33	−.30
			Difficult synonyms				
Fastidious–finicky	.51	.13	.38	Imperturbable–nonchalant	.14	.27	−.13
Testy–touchy	.66	.32	.34	Brusque–curt	.33	.45	−.12
			Easy antonyms				
Conscientious–unscrupulous	−.55	−.16	.39	Humble–proud	−.05	−.32	−.27
Deliberate–impulsive	−.48	−.10	.38	Indirect–straightforward	−.23	−.48	−.25
			Difficult antonyms				
Affable–reserved	−.49	.12	.61				
Obtrusive–shy	−.62	−.08	.54				

Note. DEF = definition condition; NON = nondefinition condition; D − N = difference. All values are based on the continuous (1–9) response scale and the correlation coefficient as the index of consistency. The correlations were computed across the subjects for each synonym and antonym pair.

pair (and in an antonym pair after the response values for one term of the pair have been reflected about the midpoint of the rating scale) is, the greater is the semantic consistency elicited by that pair; (b) the more extreme the mean values from the midpoint

Table 6
Prediction of Interpair Differences in Semantic Consistency: Correlations Between Consistency and Each Potential Predictor

	Synonyms		Antonyms[a]		All Pairs	
Potential predictor	DEF	NON	DEF	NON	DEF	NON
			Associated with consistency			
Squared difference between *Ms*	−.40**	−.18	−.21	−.42**	−.29**	−.43**
Squared difference of *Ms* from middle of response scale	.33*	.38*	.16	.37*	.24*	.31**
Multiple *R*	.46	.40	.24	.48	.33	.48
			Not associated with consistency			
Squared difference between *Vars*	−.07	−.12	−.20	.18	−.19	.01
Average of *Vars*	.08	−.13	.09	−.05	.05	−.08
Average of *Ms*	−.06	.18	.43**	.08	.15	.12

Note. DEF = definition condition; NON = nondefinition condition. For synonym pairs and antonym pairs each, n = 40 (hence for all pairs, N = 80). The semantic consistency values are based on the continuous (1–9) response scale and the correlation coefficient (r) as the index of consistency.
[a] For the antonyms, all indexes were calculated after reflection of the response values for one term in each pair; consequently, the correlations should be of the same sign for the antonyms as for the synonyms.
* $p < .05$. ** $p < .01$.

of the response scale, the greater is the consistency of the pair.

In the upper part of Table 6 we show that these expectations were confirmed in the present study, but the relations were not strong. The multiple correlations (uncorrected for capitalization on chance) of these two indexes with semantic consistency ranged from .24 to .48, indicating that most of the variance in the consistency values cannot be explained by these itemmetric characteristics. Nor did the three other indexes included in Table 6 add any significant variance to the prediction of interpair consistency. Clearly, other factors than these are affecting the interpair consistency values. A crucial problem for future research is the specification of these missing factors.

Analyses of Individual Differences

When one turns from analyses of interpair differences to analyses of individual subjects, the findings change quite dramatically. In Table 3 we show the effect of providing definitions on the consistency of both the mean and the median individual in each condition, for each type of pair, thereby permitting the reader to compare the corresponding values for subjects and for pairs of synonyms and antonyms. For every type of pair and in each of the two conditions, the *range* of consistency values was larger for subjects than for pairs, usually substantially so. This finding, the opposite of that found in our previous studies, reflects the fact that the present sample was more heterogeneous in education, verbal intelligence, and age than were all past samples of college students, and of course that here there was no culling of subjects on the basis of their semantic consistency.

An alternative method for measuring the effect of definitions on the semantic consistency of individual subjects is to examine the point–biserial correlations between the experimental condition of the subject (nondefinition = 1; definition = 2) and his or her consistency. These correlations are displayed in Table 7, for two types of consistency indexes (r and %), and for each of the four types of pairs, for all synonyms, for all antonyms, for all easy synonyms and antonyms, for all difficult synonyms and antonyms, and for all 80 pairs. The correlations in Table 7 show that one quarter or more of the variance in the consistency values can be attributed to the provision of definitions when these values are aggregated across the difficult 40 pairs or across all 80 pairs. Clearly, definitions do make a difference—in this case, quite a substantial one.

Table 7
Effect of Providing Definitions on Semantic Consistency of Individual Subjects: Point-Biserial Correlations Between Consistency and Experimental Condition (NON vs. DEF)

Type of pair	Consistency index	
	r	%
All 80 pairs	.53	.47
Difficult 40 pairs	.58	.51
Easy 40 pairs	.31	.25
All 40 antonyms	.50	.39
All 40 synonyms	.47	.38
Difficult antonyms	.58	.44
Difficult synonyms	.46	.42
Easy antonyms	.22	.23
Easy synonyms	.27	.09

Note. $N = 160$. DEF = definition condition; NON = nondefinition condition.

The effects of definitions shown in Table 7 are, in all but one case, larger for the r than for the % index of consistency. However, the rank–order correlation between the two sets of values was .92. In Table 8 are the correlations between the various sets of synonym and antonym pairs, given separately for each index of consistency (r and %). Across-set generality was higher with the r index than with %, higher in the definition than in the nondefinition conditions, and of course higher when the consistency values are aggregated across 40 pairs (all synonyms vs. all antonyms or all easy vs. all difficult pairs) than across only 20 pairs.

In Table 9 we present the correlations between each of the two consistency indexes (r and %). Generality across these two indexes was extremely high, and at the highest level of aggregation (all 80 pairs) the correlation was close to unity. Combining the results from both Tables 8 and 9, these generality values suggest that, to be measured most

Table 8
Generality of Individual Differences in Semantic Consistency Across Types of Pairs

Type of pair	Consistency index			
	r		%	
	DEF	NON	DEF	NON
All synonyms vs. all antonyms	.61	.50	.14	.26
Easy pairs vs. difficult pairs	.68	.22	.48	.24
Difficult synonyms vs. difficult antonyms	.53	.41	.25	.34
Difficult synonyms vs. easy antonyms	.49	.32	.18	.22
Difficult antonyms vs. easy antonyms	.52	.09	.58	.29
Difficult antonyms vs. easy synonyms	.50	.08	−.02	.00
Easy synonyms vs. easy antonyms	.25	.20	−.04	−.08
Difficult synonyms vs. easy synonyms	.37	.05	.17	−.04

Note. DEF = definition condition; NON = nondefinition condition.

reliably, individual differences in semantic consistency should be measured with the correlational index and aggregated across as many pairs as possible. As a consequence, in our attempts to predict individual differences in semantic consistency, we report the findings that are based on the correlational index aggregated across all 80 pairs.

Predicting Individual Differences in Semantic Consistency

It has been repeatedly declared of late that psychometric instruments and especially scale scores from personality inventories are not capable of predicting any sort of individual differences with correlations above .30. To the extent that this aphorism is now accepted, the correlations presented in Table 10 should be marvelously surprising. In the upper section of this table are the correlations between semantic consistency and each of eight CPI scales, the measure of intelligence (Terman CMT), the subject's educational level, and his or her sex (male = 0, female = 1). The left half of Table 11 shows the results that occurred when this set of predictors was used to predict semantic consistency in stepwise regression analyses.

Table 9
*Generality of Individual Differences in Semantic Consistency Across Two Indexes (*r *vs. %)*

Type of pair	Condition	
	DEF	NON
All 80 pairs	.93	.88
40 difficult pairs	.88	.89
40 easy pairs	.87	.76
40 antonyms	.89	.77
40 synonyms	.78	.73
20 easy antonyms	.89	.79
20 difficult antonyms	.80	.73
20 difficult synonyms	.78	.81
20 easy synonyms	.72	.54

Note. DEF = definition condition; NON = nondefinition condition.

To digest these findings most efficiently, let us begin with the regression analyses for all 160 subjects, in the bottom left section of Table 11. The impact of a situational influence (the provision of definitions), which accounted for over one quarter of the total variance in semantic consistency, was roughly doubled by the inclusion of a few measures of individual differences. That is, by the addition of information about a subject's sex and the inclusion of the Intellectual Efficiency (Ie) scale score from the CPI, approximately 50% of the variance in individual differences in semantic consistency can be predicted. Clearly, semantic consistency is a joint function of the experimental conditions and the characteristics of the individual subject.

Before we discuss those individual differences that were most strongly related to consistency, it is important to mention two that were not so related, the age of the subject and his or her score on Snyder's (1974) Self-Monitoring scale. Those measures that did

predict semantic consistency cluster into six categories (see Table 10): (a) verbal intelligence (from the CMT) and education (whose intercorrelation was .53 in this sample); (b) the cognitive triad of CPI scales (Ac, Ai, and Ie), whose average intercorrelation was roughly .60; (c) Well-Being (Wb); (d) Responsibility (Re); (e) Self-Acceptance (Sa) and Dominance (Do), which correlated .66; and (f) the sex of the subject and its CPI surrogate, Femininity (Fe), which correlated .63. Of these six predictor sets, the first five were strongly related to semantic consistency in the nondefinition condition, whereas sex was not. On the other hand, the sex of the subject (specifically being female rather than male) and the CPI Well-Being (Wb) scale were both correlated over .40 with semantic consistency in the definition condition.[1]

Table 10
Major Predictors of Semantic Consistency in Each Condition

	Condition	
	Non-definition	Definition
Single predictors		
CPI: Ai	.55**	.30**
CPI: Ie	.54**	.34**
CPI: Ac	.52**	.34**
CPI: Do	.52**	.13
CPI: Sa	.48**	.19
API: Re	.48**	.24*
CPI: Wb	.38**	.43**
CPI: Fe	.07	.25*
Sex (Female)	.19	.44**
CMT	.52**	.21
Education	.45**	.27*
Predictor composites		
Ai + Do + Re	.71**	.29**
Wb + Fe + Sa	.49**	.52**
CMT + education	.56**	.27*
Unrotated CPI components		
Factor I	.65**	.41**
Factor II	−.10	.05
Factor III	.13	.07
Factor IV	.09	.16
Varimax-rotated CPI components		
Factor 1	.49**	.21
Factor 2	.25*	.16
Factor 3	.25*	.24*
Factor 4	.28*	.29**

Note. CPI = California Psychological Inventory; Ai = Achievement via Independence; Ie = Intellectual Efficiency; Ac = Achievement via Conformance; Do = Dominance; Sa = Self-Acceptance; Re = Responsibility; Wb = Well-Being; Fe = Femininity; CMT = Terman's Concept Mastery Test. The consistency values are based on the continuous (1–9) response scale, with the correlational (r) index, across all 80 synonym and antonym pairs. In each condition, $n = 80$.
* $p < .05$. ** $p < .01$.

To aid in the interpretation of the CPI-based predictors, we factor analyzed the 20 CPI scales that were used in this study, using the principal component method of factor extraction, followed by a varimax rotation of the first four components.[2] The correlations between semantic consistency and each of the first four unrotated, and each of the four rotated, components are presented at the bottom of Table 10. Surprisingly, semantic consistency was highly related to the first unrotated component ($r = .65$ in the nondefinition and .41 in the definition conditions), and completely unrelated to any of the other

[1] In addition, we constructed two aggregate CPI predictors: Ai + Do + Re, the CPI scales most powerfully associated with consistency in the nondefinition condition, and Wb + Fe + Sa, the CPI scales most strongly associated with consistency in the definition condition. The latter composite turned out to be a general predictor of semantic consistency in both experimental conditions.

How well are these unit-weighted CPI composites predictive of consistency relative to the optimal (and thus chance-inflated) composites available within the total set of potential predictors? In the nondefinition condition, the use of Ai, Do, and CMT with optimal regression weights produced a multiple correlation of .76 (see Table 11). In comparison, the unit-weighted CPI coomposite, Ai + Do + Re, was correlated .71. In the definition condition, the optimal-weighted linear combination (sex, Wb, and education) had a multiple correlation of .58 with consistency. In comparison, the unit-weighted CPI composite (Wb + Fe + Sa) correlated .52. (In the nondefinition condition, this latter CPI composite correlated .49).

Overall, then, one can isolate three major classes of consistency predictors: (a) the experimental condition (definition vs. nondefinition), (b) a simple CPI-based personality composite (Wb + Fe + Sa), and (c) the intelligence (CMT) and education combination. In the total sample of 160 subjects, the most potent of these three classes was condition ($r = .53$); the addition of Wb + Fe + Sa raised the multiple correlation to .69; and the inclusion of education and CMT scores raised the multiple correlation to .73.

[2] We repeated these analyses using a variety of methods of factor extraction and rotation, with no substantial changes in the findings.

Table 11
Prediction of Individual Differences in Semantic Consistency: Results of Some Stepwise Regression Analyses

Condition and step	Individual CPI Scales				Unrotated CPI Components			
	Predictor	*R*	*p*	*r*	Predictor	*R*	*p*	*r*
NON								
1	CPI: Ai	.55	.000	.55	CPI: Factor I	.65	.000	.65
2	CPI: Do	.73	.000	.52	CMT	.72	.000	.52
3	CMT	.76	.003	.52	Education	.73	.047	.45
DEF								
1	Sex	.44	.000	.44	Sex	.44	.000	.44
2	CPI: Wb	.54	.002	.43	CPI: Factor I	.53	.004	.41
3	Education	.58	.037	.27	Education	.54	.160	.27
Both								
1	Condition	.53	.000	.53	Condition	.53	.000	.53
2	CPI: Ie	.65	.000	.36	CPI: Factor I	.69	.000	.44
3	Sex	.68	.000	.22	Sex	.72	.001	.22
4	Education	.71	.001	.30	Education	.74	.002	.30

Note. CPI = California Psychological Inventory; Ai = Achievement via Independence; Do = Dominance; Wb = Well-Being; Ie = Intellectual Efficiency; CMT = Terman's Concept Mastery Test. NON = nondefinition condition; DEF = definition condition. The consistency values are based on the continuous (1–9) response scale, with the correlational (*r*) index, across all 80 pairs. (*R* = multiple correlation; *p* = significance level; *r* = zero-order correlation)

unrotated components. As a consequence, when the components were rotated to a simple structure criterion and the variance in the unrotated first component was distributed across the other three components, each of the four resulting varimax factors was less predictive of semantic consistency than had been the first unrotated factor alone.

The results from stepwise multiple regression analyses when the first unrotated CPI component was included as a predictor (and correspondingly all individual CPI scales were omitted) are presented in the right half of Table 11. A comparison of the resulting multiple correlations with those in the left half of the table shows that the substitution of the first unrotated CPI component for the individual CPI scale scores did not reduce predictability substantially in either of the two experimental conditions, in spite of the fact that the multiple correlations that are based on an initial pool of 20 CPI scales are far more likely to be chance inflated than are those based solely on one unrotated component. Moreover, for the complete sample with experimental condition as a predictor variable, the multiple correlations that are based on the first unrotated CPI component actually turned out to be higher than those based on the complete predictor set.

These findings suggest that the major source of the high CPI-based predictions of individual differences in semantic consistency resides in the fact that variance that is common to subjects' scores on all of the CPI scales. Because each of the CPI scales is scored in the socially desirable direction, and because all of the scales loaded positively on the first unrotated component,[3] some highly general response determinant is implicated. The inevitably entwined constructs of general adjustment or social desirability response set appear to be the most likely candidates for this stellar role.

Predicting Semantic Consistency From Responses to the 191 Trait Terms

Are there particular traits (e.g., *accurate, careful, conscientious, precise, reflective*) that subjects of high semantic consistency view as more self-descriptive than do those of lower

[3] The loadings of each of the 20 CPI scales on the first unrotated component were as follows: Ie, .86; To, .82; Wb, .79; Ac, .75; Ai, .74; Cs, .74; Py, .73; Sy, .66; AUT, .64; Re, .60; Sp, .60; Do, .58; EMP, .58; Gi, .53; Sa, .48; Sc, .46; So, .39; Cm, .36; Fx, .34; and Fe, .09.

consistency? To find out, we correlated the subjects' response values for each of the 191 trait-descriptive terms in the self-rating task with their overall semantic consistency averaged across all 80 synonym and antonym pairs; for these analyses, we measured consistency with the correlational index.

Interestingly, the terms that were the most highly related to semantic consistency in the definition condition differed from those in the nondefinition condition; the correlation between the two sets of correlations across all 191 terms was only .52.[4] However, in spite of these differences in the *particular* terms that were most highly related to consistency, there was a whopping similarity in the *character* of the terms that were positively so related as compared with those that were negatively so related. In both conditions, virtually all of the terms that were positively related to semantic consistency were evaluatively desirable, whereas most of the terms that were negatively related to consistency were undesirable ones. Because the mean response of the subjects is a reasonable surrogate of the social desirability of a term (see Edwards, 1957), the correlation across the 191 terms between the mean of each term and its correlation with consistency provides an estimate of the effect of the terms' desirability values on their relation with semantic consistency. These correlations were .54 in the nondefinition and .70 in the definition conditions.

What if one were to partial out the desirability values from these consistency correlations? Would the resulting patterns of residual correlations now be the same in the two conditions? This hypothesis proved to be unfounded. The correlation across the 191 terms in their residual correlations with semantic consistency (the mean values having been partialed out) was only .12. Once again, as with the findings based on the CPI scales, some highly general response determinant such as adjustment or social desirability response set has been found to be a major predictor of semantic consistency.

Discussion

The most reasonable way to discuss the variety of findings from this study is to assemble its results as they bear on the initial hypotheses described in the introduction to this report.

Response Biases

Over the years, research on response sets and styles has come to focus on three major varieties: (a) extremeness response style, (b) acquiescence response bias, and (c) social desirability response set. The first of these, individual differences in subjects' tendency to use the end values of the rating scale as compared with the middle values, has turned out to be a quite pervasive and general response style. In the present context, individual differences in extremeness should serve to increase semantic consistency to both synonym and antonym pairs. On the other hand, individual differences in acquiescence (the tendency to accept vs. reject terms as self-descriptive, regardless of their content), which has not turned out to have much in the way of across-task generality, will serve to increase semantic consistency among synonym pairs and to decrease it among antonym pairs.

The effects of standard scoring each individual's response values should serve to attenuate, if not eliminate, individual differences in both extremeness and acquiescence response biases. Indeed, if the semantic consistency of the average synonym pair is actually the same as that of the average antonym pair, but for the effects of acquiescence, then standard scoring should serve to make the two averages the same. The results presented in Table 4 show that there was a significant acquiescence effect within these self-ratings, but that the effect was fairly small. On the basis of the correlational index of semantic consistency, the average difference in consistency correlations between synonyms and antonyms, which initially ranged from .12 to .27 over the two conditions and the two difficulty levels, ranged from .06 to .12 when the response protocols were standard scored. Said another way, the difference in correlations between synonyms and antonyms was roughly cut in half by standard scoring, although even in this latter case semantic con-

[4] The correlation between the two sets of mean values across the 191 terms was .85.

sistency was still higher for synonym than for antonym pairs.

When one turns from the effects of individual differences in acquiescence to those in social desirability response set, the conceptual problems of interpretation increase enormously. Whereas it is clear that it is not necessary for subjects to respond in the socially desirable direction to achieve high semantic consistency values (for example, if a subject's responses were negatively correlated with the desirability values, his or her semantic consistency could be as high if they were positively so correlated), it is equally clear that because subjects respond in the desirable direction to most items, indexes of desirability response tendency will generally be highly related to indexes of semantic consistency.

Nor is it obvious how these two conceptually related indexes can ever be completely unconfounded. Because the definition of a good synonym includes similarity in desirability (and of a good antonym, symmetrical opposition in desirability), all tendencies to respond in the desirable direction will inevitably result in increased semantic consistency. However, the reverse is not true. A subject could respond in the desirable direction to half the terms and the undesirable direction to the other half—thus achieving a zero correlation with desirability—and still be perfectly consistent semantically, as long as the corresponding terms in each synonym and antonym pair are responded to in a semantically consistent direction. However, because relatively few subjects do not show substantial correlations between their responses and the desirabilities of the terms, it should come as no great surprise that desirable terms will tend to be positively related to consistency, and undesirable terms negatively related.

Moreover, to the extent that individual differences in adjustment (or social desirability response set) generalize across questionnaires, it also is not surprising that semantic consistency was highly related to that variance common to all of the CPI scales (here indexed by the first unrotated CPI component). Indeed, these CPI findings are remarkably similar to those found by Goldberg (1978b), who demonstrated that individual differences in test–retest consistency to CPI and Minnesota Multiphasic Personality Inventory (MMPI) items were highly related to individual differences in responding in the socially desirable direction to such items. Moreover, those CPI scales that were the most predictive of semantic consistency in the present study (e.g., Ie) were precisely the ones that were also most predictive of test–retest item consistency (see Goldberg's 1978b Tables 4 and 6).

Indeed, one could now argue that the Intellectual Efficiency (Ie) scale of the CPI might function as a near-ideal "validity" scale for culling subjects whose item responses are likely to be inconsistent both semantically and temporally. The Ie scale (which was empirically constructed with an intelligence test as a criterion measure) had the highest loading on the first unrotated CPI component (see Footnote 2), was the CPI scale most predictive of semantic consistency in the total sample of subjects (see Tables 10 and 11), and was the most potent CPI predictor of test–retest item consistency in Goldberg's (1978b) study. It appears as if the Ie scale may tap both of the major sources of variance associated with individual differences in consistency, namely verbal intelligence and the adjustment/social-desirability confound.

Word Difficulty

This hypothesis was strongly confirmed in the present study. Evidence that implicates word difficulty (and verbal comprehension) as potent factors affecting semantic consistency includes (a) the substantial impact on consistency provided by dictionary definitions; (b) the significant correlations in the nondefinition condition between individual differences in consistency and scores on the Terman CMT, as well as years of education; and (c) again in the nondefinition condition, the high correlations between individual differences in consistency and those three CPI scales that were constructed from measures of intelligence (Ie), college achievement (Ai), and high school achievement (Ac).

Because verbal intelligence has been implicated as a factor in semantic inconsistency for cases in which definitions were not provided, can it similarly be invoked to explain individual differences in consistency for cases in which they were? After all, definitions are composed of words, and some subjects may

not know the meaning of some of the defining terms. However, neither the Terman CMT nor years of education were significantly associated with most indexes of semantic consistency in the definition condition.

Perhaps some subjects did not bother to read the definitions for those terms that they thought they understood, even when they actually had idiosyncratic understandings of the terms' meanings. To the extent that the subjects read the task instructions and then responded accordingly, this becomes a less tenable hypothesis. The instructions in the definition condition read in part as follows:

> Following each adjective, a short definition is provided which describes the kind of person to whom the adjective applies. Since words mean different things to different people and since the words themselves may have several meanings, it is important that the *same meaning* be used by everyone. So, please *read both the word and its definition* before making your rating, *even for easy words.* (Emphasis in original instructions.)

On the other hand, it is possible that one might invoke individual differences in task compliance as the link between the sex of the subject and semantic consistency in the definition condition. Otherwise, it is difficult to understand why female subjects were so much more consistent than male subjects in the definition condition yet not in the nondefinition condition.

Implications and Further Hypotheses

In summary, the results of this study have some important implications. Although subjects can be induced to give self-ratings under many formats, the meaning of these responses is ambiguous under the standard one (Loehlin, 1961, 1967). One cannot assume that such ratings reflect the subject's accurate perception of the dictionary definitions of the stimulus terms. Unless definitions are provided, it is difficult to know precisely what a response signifies, and this seems to be true for all but the most simple trait adjectives. Indeed, one of the most surprising aspects of this study was the difference found between sets of "difficult" and "easy" pairs, given the nature of the adjectives included in the "difficult" set (e.g., *unruly–docile, decisive–irresolute,* and *cautious–wary*).

The difference in semantic consistency between antonym and synonym sets found in previous studies—responses to antonym pairs are typically less consistent than those to synonym pairs—was again replicated in this study. One explanation for inconsistent responses to antonym pairs derives from the nature of the trait construct. It is generally assumed that trait attributions are averaged summaries of a person's past behaviors (Hampshire, 1953). Thus persons' ratings of their generosity may be a function of the percentage of occasions on which they have acted generously, relative to their opportunities for generous behavior. If one usually contributes time or money to worthy causes, one may feel justified labeling oneself generous regardless of the size of the gifts. However, the intensity of one's behaviors might also be a basis for trait ascriptions (Goldberg, 1981). If so, then it is neither inconsistent nor contradictory to assert that one is both generous and stingy, because one may sometimes be unusually generous and at different times be unusually stingy.

Moreover, when responding to the two terms in an antonym pair, subjects may average over different behavioral domains; that is, subjects may think of one set of behavioral contexts in producing a rating for one term and a second set for its opposite. One who is dependable in academic matters may look to these behaviors in rating onself as *conscientious;* the same person might be slovenly in personal affairs and rate himself or herself as *negligent* on these grounds. Inspecting these responses from an external frame of reference and being unadvised of the different behavioral contexts for the responses, one would term them "inconsistent," because it seems illogical for a person to be both conscientious and negligent. Because the averaging has taken place over different behavioral domains, however, the subject justifiably may not experience these self-descriptions as inconsistent.

Finally, it is clear from the results in the definition condition that definitions per se provide no guarantee that all synonym and antonym pairs will be rated in a semantically consistent manner. Just as research with public opinion polls has revealed that seemingly minor differences in the wording of a question can have enormous effects on the average

response to it, so seemingly small differences in definitions (as found in different dictionaries) can probably produce substantial changes in the semantic consistency values for any individual synonym and antonym pair. It has long been known that trait-descriptive terms function in our lexicons like "fuzzy sets." Perhaps the major implication of the results of this study is that such terms are even more fuzzy than we have ever thought.

References

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs, 47* (1, Whole No. 211).

Bentler, P. M., Jackson, D. N., & Messick, S. (1971). The identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin, 76,* 186–204.

Cattell, R. B. (1957). *Personality and motivation structure and measurement.* New York: World Book.

Cohen, J. (1969). r_c: A profile similarity coefficient invariant over variable reflection. *Psychological Bulletin, 71,* 281–284.

D'Andrade, R. G. (1965). Trait psychology and componential analysis. *American Anthropologist, 67,* 215–228.

D'Andrade, R. G. (1974). Memory and the assessment of behavior. In H. M. Blalock, Jr. (Ed.), *Measurement in the social sciences: Theories and strategies* (pp. 159–186). Chicago: Aldine.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research.* New York: Dryden.

Goldberg, L. R. (1978a, March). *Language and personality: Developing a taxonomy of personality-descriptive terms (A progress report and research proposal).* Institute for the Measurement of Personality, Eugene, Oregon.

Goldberg, L. R. (1978b). The reliability of reliability: The generality and correlates of intra-individual consistency in responses to structured personality inventories. *Applied Psychological Measurement, 2,* 269–291.

Goldberg, L. R. (1981). Unconfounding situational attributions from uncertain, neutral, and ambiguous ones: A psychometric analysis of descriptions of oneself and various types of others. *Journal of Personality and Social Psychology, 41,* 517–552.

Goldberg, L. R. (1982). From Ace to Zombie: Some explorations in the language of personality. In C. D. Spielberger & J. N. Butcher (Eds.), *Advances in personality assessment: Volume I* (pp. 203–234). Hillsdale, NJ: Erlbaum.

Hampshire, S. (1953). Dispositions. *Analysis, 14,* 5–11.

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology, 33,* 307–316.

Kurtines, W. (1978). A measure of autonomy. *Journal of Personality Assessment, 42,* 253–257.

Levy, L. H., & Dugan, R. D. (1960). A constant error approach to the study of dimensions of social perception. *Journal of Abnormal and Social Psychology, 61,* 21–24.

Loehlin, J. C. (1961). Word meanings and self-descriptions. *Journal of Abnormal and Social Psychology, 62,* 28–34.

Loehlin, J. C. (1967). Word meanings and self-descriptions: A replication and extension. *Journal of Personality and Social Psychology, 5,* 107–110.

Mulaik, S. A. (1964). Are personality factors raters' conceptual factors? *Journal of Consulting Psychology, 28,* 506–511.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology, 66,* 574–583.

Norman, W. T. (1967, April). *2800 personality trait descriptors: Normative operating characteristics for a university population.* Technical report, Department of Psychology, University of Michigan.

Norman, W. T., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology, 4,* 681–691.

Shweder, R. A. (1975). How relevant is an individual difference theory of personality? *Journal of Personality, 43,* 455–484.

Shweder, R. A. (1977). Likeness and likelihood in everyday thought: Magical thinking in judgments about personality. *Current Anthropology, 18,* 637–658.

Shweder, R. A. (1979). Rethinking culture and personality theory. Part I: A critical examination of two classical postulates. *Ethos, 7,* 255–278.

Shweder, R. A. (1982). Fact and artifact in trait perception: The systematic distortion hypothesis. In B. A. Maher & W. B. Maher (Eds.), *Progress in experimental personality research* (Vol. 11, pp. 65–100). New York: Academic Press.

Shweder, R. A., & D'Andrade, R. G. (1980). The systematic distortion hypothesis. In R. A. Shweder (Ed.), *New directions for methodology of social and behavioral science: Number 4. Fallible judgment in behavioral research* (pp. 37–58). San Francisco: Jossey-Bass.

Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology, 30,* 526–537.

Terman, L. (1956). *Manual for the Concept Mastery Test, Form T.* New York: Psychological Corporation.

Received February 14, 1983
Revision received August 15, 1983 ■