Social Cognition, Vol. 5, No. 1, 1987, pp. 26–57

WHEN IS A SCHEMA NOT A SCHEMA? THE "BIG FIVE" TRAITS AS COGNITIVE STRUCTURES

STEVENS S. SMITH AND JOHN F. KIHLSTROM
University of Wisconsin–Madison

Five studies investigated personality trait structures as cognitive schemas. Norman's (1963) "Big Five" trait dimensions were hypothesized to function as schemas to organize the lexicon of subordinate traits. Experiments 1 and 2 demonstrated that subjects employ an implicit theory of personality similar to the "Big Five" structure in making judgments about trait co-occurrence and semantic similarity. Experiments 3 and 4, however, failed to show clustering by trait category in a free-recall paradigm, regardless of whether the memory task was framed in a social context or not. Experiment 5 demonstrated considerable overlap in meaning in the subordinate traits in the Norman list, suggesting that the results in Experiments 3 and 4 may reflect a lack of distinctiveness in the trait–factor relationships. The knowledge stuctures comprising implicit personality theory are not automatically activated by schema-relevant stimulus information.

Contemporary conceptions of persons as active processors of information underscore the role of cognitive structures in the acquisition and organization of both social and nonsocial knowledge. Confronted by complex, diverse, and fragmentary stimuli in the environment, the person must somehow organize the available information in an efficient and economical manner in order to make use of it. Comprehension involves a dynamic process in which an individual imposes structure on incoming information through selective attention and

This research was conducted as part of a master's thesis submitted by Stevens S. Smith to the Department of Psychology, University of Wisconsin–Madison, and was supported in part by Grant No. MH-35856 from the National Institute of Mental Health. Committee members Jean Chapman and Lola Lopes provided helpful comments on an earlier version of this manuscript. We also wish to thank Del Paulhus, Carol Martin, and Jerry Wiggins for their comments on the studies. We gratefully acknowledge the assistance of John Grych in Experiment 4. Requests for reprints should be sent to Stevens S. Smith, Department of Psychology, University of Wisconsin, W. J. Brogden Psychology Building, 1202 West Johnson Street, Madison, WI 53706.

categorization. The individual then goes beyond the information given by the stimulus, filling in the gaps with information derived from memory, expectations, or inferences. The cognitive structures that mediate this organizational and inferential activity are known collectively as "schemas" (Bartlett, 1932; Neisser, 1967, 1976). More than 50 years after they were introduced to cognitive psychology, there remains considerable controversy about the precise nature and function of cognitive schemas (e.g., Brewer & Nakamura, 1984). Nevertheless, it is generally agreed that schemas are structures for representing large sets of organized information about concepts, and serve to facilitate inference concerning these concepts and specific instances of them (Anderson, 1985; Rumelhart, 1980).

Currently, personality and social psychologists are devoting considerable effort to understanding the schemas that guide social cognition (Hastie, 1980; Taylor & Crocker, 1981). Historically, the study of social schemas appears to have begun with work on implicit personality theory (Bruner & Tagiuri, 1954; Schneider, 1973)—the "kinds of naive, implicit 'theories' of personality . . . people work with when they form an impression of another" (Bruner & Tagiuri, 1954, p. 649). In part, implicit personality theory consists of people's intuitive notions concerning the relations among features of personality. For example, people appear to possess structured sets of beliefs or expectations about the covariations to be observed among various behaviors and traits. We expect that people who tend to smile will also tend to laugh, and that those who are careless will also be undependable. In part, then, implicit personality theory can be represented as a conceptual hierarchy with specific behaviors (e.g., "smiling" and "laughing") grouped under somewhat more abstract trait concepts (e.g., "happy"). Similarly, low-level or "primary" traits (e.g., "careless" and "undependable") can be grouped under even more abstract "secondary" trait concepts (e.g., "conscientiousness"). Such a hierarchy provides a cognitive basis for making inferences based on certain observations (e.g., that people who smile also laugh, that people who are talkative are also sociable), as well as for grouping certain observations and inferences together (i.e., that laughing and smiling, or talkativeness and sociability, have something in common).

The perceived relations among personality features have been investigated in several ways, such as simple correlation, factor analysis, cluster analysis, and multidimensional scaling (for reviews, see Kim & Rosenberg, 1980; Rosenberg & Sedlak, 1972; Schneider, 1973). At a fairly abstract level, similarity ratings appear to reveal two major dimensions, intellectual and social, running through the domain of personality traits (Rosenberg, Nelson, & Vivekananthan, 1968). These

dimensions are correlated, so that at the most abstract level the trait lexicon (and personality impressions) appears to be mediated by a single dimension of evaluation (Osgood, Suci, & Tannenbaum, 1957; Rosenberg, 1977).

A somewhat different view of implicit personality theory is offered by research on self-report and peer ratings. Using a set of adjectives derived from Allport and Odbert (1936) and Cattell (1947, 1957), several investigators have sought to determine the relations among the primary traits of personality. Multivariate analyses by a number of different investigators have shown that approximately five orthogonal factors account for the majority of variance in trait ratings (Digman & Inouye, 1986; Digman & Takemoto-Chock, 1981; Fiske, 1949; McCrae & Costa, 1985a, 1985b; Norman, 1963; Tupes & Christal, 1961): "Extraversion," "Agreeableness," "Conscientiousness," "Emotional Stability," and "Culture." These five factors appear with such regularity in adjective checklist ratings that Goldberg (1981) christened them the "Big Five" and suggested that they constitute a universally applicable structure suitable for representing individual differences in personality.

In traditional psychometric theories of personality, the "Big Five" (or something very much like them; see Eysenck & Eysenck, 1985; Hogan, 1983) are held to reside in the individual as dispositions to behave in various ways. But it is also likely that they form part of implicit personality theory, as linguistic categories used in organizing impressions of ourselves and other people. For example, the "accurate-reflection" hypothesis states that our intuitions about personality structure are derived from observations of co-occurrences among behaviors and traits in the real world (Block, Weiss, & Thorne, 1979; Buss & Craik, 1985; Jackson, 1982; Shweder, 1982; Schweder & D'Andrade, 1979, 1980; Weiss & Mendelsohn, 1986). This hypothesis is congruent with contemporary realist approaches to categorization, which assume that the structure of natural categories, far from being arbitrary, closely matches the structure of the world that is subjected to classification (Kripke, 1972; Putnam, 1975; Rosch, 1978). In other words, if "talkativeness" and "frankness" are empirically correlated, then they ought to be linked together as instances (or expressions) of a more general dispositional category, such as Extraversion. Therefore, we should expect to find the actual structure of personality in the world outside the mind mirrored in our mental representations of that structure.

Of course, the accurate-reflection hypothesis has come under attack from proponents of the "systematic-distortion" hypothesis, which holds that our preconceptions of personality structure influence the trait relations that we perceive in the world (e.g., Schweder

& D'Andrade, 1979, 1980). According to this view, the "Big Five" have little or no external validity, but rather reflect rater biases similar to Thorndike's (1920) "halo effect" or Newcomb's (1931) "logical error." However, it is not necessary to take a position on accurate reflection versus systematic distortion in order to determine whether the "Big Five" serve as schemas for person perception. In fact, Weiss and Mendelsohn (1986) have offered evidence suggesting that the "Big Five" are not artifacts of systematic distortion. Nevertheless, according to both hypotheses, the "Big Five" are good candidates for cognitive structures.

Several lines of research have provided evidence that the "Big Five" may function as schemas guiding impression formation. In a now-classic study, Passini and Norman (1966) obtained this five-factor structure from trait ratings of complete strangers. Because the judges had no opportunity to observe any behavior of the targets, much less co-occurrences of behavior, the factors must have been artifacts of the judges' preconceptions of the structure of personality. More recently, Cantor and Mischel (1977, 1979a, 1979b) used Norman's (1963) five factors as the basis for investigations of social categorization. They found that such categories as "extraverted person," "emotionally unstable person," and "cultured person" influenced performance in a wide variety of tasks involving person memory and impression formation. Also, Hampson and her colleagues (Hampson, John, & Goldberg, 1985) have recently extended Cantor's analyses, suggesting that the "Big Five," construed as categories rather than factors, may be somehow basic to person perception.

The experiments reported here further explored the status of the "Big Five" as cognitive structures. All of the experiments were constructed so that the subjects had the opportunity to employ the "Big Five," if in fact they were cognitively available for use—that is, if they existed as schemas or mental structures. In Experiment 1, a conceptual-similarity procedure was used to test the influence of the "Big Five" on perceptions of co-occurrences. Experiment 2 was a conceptual replication of Experiment 1, involving an illusory-correlation procedure derived from Chapman and Chapman (1969). Experiment 3 used a free-recall procedure in which the "Big Five" could serve as a rubric for organizing the list items. Experiment 4 was a conceptual replication of Experiment 3, in which the memory task was embedded in a somewhat more social context. Experiment 5 employed conceptual-similarity judgments of a sort employed by Rosenberg and Sedlak (1972) in an attempt to resolve discrepancies between the first and second pairs of studies, and to address the question posed in the title of this paper.

EXPERIMENT 1

While experiments by Norman (Norman, 1963) and others have provided good evidence for the "Big Five" using judges' ratings, other work by Passini and Norman (1966) and Hakel (1969) provided support for the internal structure of the "Big Five" in regard to implicative meaning (Wiggins, 1973)—that is, the extent to which traits are interchangeable. In the first experiment, the structure of the "Big Five" was investigated by asking subjects to make judgments of perceived co-occurrence between traits and factors (dimensions) of the "Big Five." In this procedure, subjects were instructed to choose two traits (from a list of 40 based on Norman, 1963) that they thought were descriptive of people who score low or high on a personality test (factors of the "Big Five"). It was hypothesized that subjects' judgments would reflect the "Big Five" structure.

METHOD

Subjects

A total of 40 subjects from introductory psychology classes at the University of Wisconsin–Madison volunteered to participate, and received points toward an extra-credit option.

Stimulus Materials

Table 1 shows the five personality factors identified by Norman (1963) and their constituent traits. Each factor has two poles, positive and negative, and each factor–pole combination has a set of four trait adjectives associated with it. A randomized listing of the 40 trait adjectives from Norman (1963) was printed at the top of a page, followed by a set of instructions: for example, "Certain traits may be more descriptive of people who score high on a test of Extraversion. Choose from the above list the two traits that you think are most often associated with a high score on this particular test." There were 10 such pages, corresponding to all possible combinations of the "Big Five" factors ("personality tests") and two poles ("high" or "low" scores on the ostensible personality tests). Each page contained a separate randomization of the 40 traits, and each subject completed a questionnaire with a different randomization of the 10 pages.

TABLE 1
Trait and Personality Factor Relationships from Norman (1963)

	TRAITS	
PERSONALITY FACTOR	POSITIVE POLE	NEGATIVE POLE
Extraversion	Talkative	Silent
	Frank	Secretive
	Adventurous	Cautious
	Sociable	Reclusive
Agreeableness	Good-natured	Irritable
	Not jealous	Jealous
	Gentle	Headstrong
	Cooperative	Negativistic
Conscientiousness	Tidy	Careless
	Responsible	Undependable
	Scrupulous	Unscrupulous
	Persevering	Quitting
Emotional Stability	Poised	Nervous
	Calm	Anxious
	Composed	Excitable
	Not hypochondriacal	Hypochondriacal
Culture	Artistically sensitive	Artistically insensitive
	Intellectual	Unreflective
	Refined	Crude
	Imaginative	Simple

Procedure

The experiment was conducted in group sessions in a small classroom. After written informed consent was obtained from each participant, each subject was given a questionnaire and asked to read through the instructions. The subjects were told that the experiment was concerned with their judgments about personality and that the task consisted of examining a list of traits and choosing two traits that are more descriptive of people who score low (or high) on a given personality test. As each subject finished, he or she was individually thanked for participating and given extra-credit points.

RESULTS

For each combination of factor and pole, the subjects' responses were independently coded in terms of whether the chosen trait came from (1) that factor and (2) that pole, as these relationships were deter-

mined by Norman (1963). For each of the 10 factor–pole combinations, those choices represented the four cells of a 2×2 contingency table (same vs. different factor and same vs. different pole). If the null hypothesis of equal probability of any of the 40 traits falling into the four cells were true, then the expected cell frequencies would be as follows (given a sample of 40 subjects): (1) same factor–same pole (SFSP) = 4, (2) same factor–different pole (SFDP) = 16, (3) different factor–same pole (DFSP) = 4, and (4) different factor–different pole (DFDP) = 16. The comparison of interest was SFSP (expected cell frequency = 4) versus SFDP, DFSP, and DFDP combined (expected cell frequency = 16 + 4 + 16 = 36). This comparison involved analyzing the difference between two proportions: the proportion of SFSP responses expected under a random model versus the observed proportion of SFSP responses. The difference between these two proportions was tested by a large-sample approximation to the Fisher exact test (Marascuilo & McSweeney, 1977). Subjects' first and second choices were analyzed separately.

As shown in Table 2, the results of the analyses for the first and second choices showed strong evidence for subjects' perception of

TABLE 2
Frequencies and Fisher *z* Values for First and Second Choices in Experiment 1

	CELL FREQUENCIES: FIRST CHOICE			CELL FREQUENCIES: SECOND CHOICE		
FACTOR	SPSF	ALL OTHER	*z*	SFSP	ALL OTHER	*z*
Extraversion						
High	27	13	5.28**	21	19	4.10**
Low	31	9	6.09**	24	16	4.69**
Agreeableness						
High	28	12	5.48**	20	20	3.90**
Low	30	10	5.88**	22	18	4.30**
Conscientiousness						
High	22	18	4.30**	8	32	2.14*
Low	23	17	4.49**	24	16	4.69**
Emotional Stability						
High	23	17	4.49**	19	21	3.71**
Low	25	15	4.88**	20	20	3.90**
Culture						
High	28	12	5.48**	19	21	3.71**
Low	25	15	4.88**	20	20	3.90**

*$p < .05$.

**$p < .001$.

some kind of trait relationships. All 10 factor–pole combinations for the first choice showed departures from the expected cell proportions in the tabulated data (all p's $< .05$). Similar results were found for the second choice, with the exception of the high Conscientiousness factor–pole combination.

EXPERIMENT 2

The results in Experiment 1 strongly suggested that subjects used their implicit theories of personality in making judgments about which traits are associated with target persons described as scoring low or high on a given personality test. In Experiment 2, subjects were asked to form impressions of personality–trait relations based on information given in a series of stimulus presentations. The basic method used here was similar to the method used by Chapman and Chapman (1969) to investigate illusory correlation. The objective was to determine whether subjects would be biased in their perception of the relationship between personality factors and subordinate traits, and whether such a bias, if indeed it existed, corresponded to the structure obtained by Norman (1963).

METHOD

Subjects

A total of 20 students from introductory psychology classes at the University of Wisconsin–Madison served as subjects in the present experiment. In return for their participation, subjects received points toward an extra-credit option.

Stimulus Materials

The trait adjectives from Norman (1963) were used to construct stimuli consisting of (1) the initials of a hypothetical person; (2) a trait that was ostensibly derived from peer descriptions of the person; and (3) whether the person scored low or high on two ostensible personality tests, as in the following example: "KLM is described by friends as good-natured and scores high on tests of Agreeableness and Extraversion." Stimuli were constructed such that there were 400 different presentations consisting of combinations of the 40 traits, five factors (represented as "personality tests"), and two poles (represented as

"high" and "low" scores on the personality tests) from Norman (1963).

Combinations of traits, scores, and personality traits were constructed according to a computer algorithm that combined each trait with either a high or a low score on both of two personality tests. The algorithm was designed such that each of the 40 traits appeared 10 times, each type of score (low or high) appeared 200 times, and each type of "test" appeared 40 times. In each presentation, "tests" (or factors) were presented in pairs (as in the example above), such that each of the five factors appeared as the first member of the pair half the time and as the second member of the pair half the time. The algorithm combined traits, scores, and tests such that each trait was paired with a high and a low score on each of the five personality tests. Thus, within the set of 400 presentations, there was no actual relationship between the scores on the personality tests and the trait descriptions. The computer program also generated a unique set of three initials (representing a hypothetical person) for each of the 400 presentations. Thus, the 400 presentations was designed to correspond to 400 different people. Each presentation was displayed on a computer screen for 5 seconds for a total time of approximately 40 minutes for the complete set of 400 stimuli. Each subject saw a uniquely randomized ordering of the 400 presentations, thus controlling for order effects.

Procedure

Subjects were tested individually during an hour-long session. After explaining the basic procedure and obtaining informed consent, the experimenter initiated the computer program, worked though the instructions (displayed on the computer screen) with each subject, and familiarized the subject with the use of the keyboard. The subject was informed that the experiment was concerned with how people form impressions of personality and that the presentations contained information about the personalities of many different people, identified only by their initials (e.g., "KLM"). The subject then viewed a sample presentation. The experimenter explained that subsequent presentations would be similar but that each presentation represented a different person. The subject was instructed to form a general impression of the relationships among the traits and personality tests rather than focusing on scores of particular individuals.

After viewing the 400 stimuli, a randomized list of the 40 traits was displayed at the top of the computer screen, and the subject was asked to choose those two traits from the list that were associated

most often with a particular score (low or high) on one of the five personality tests. Altogether, each subject made choices for all 10 factor–pole combinations, each from a separate randomized list of the 40 traits. Upon completion of this phase, the subject was debriefed, given extra-credit points, and dismissed.

RESULTS

Subjects' responses on the judgmental tasks in this experiment could be categorized in the same way as subject responses in Experiment 1. Accordingly, the same procedure for categorizing subjects' choices was used to generate expected and observed proportions for SFSP responses on the 10 factor–pole combinations. Expected frequencies for Experiment 2 corresponded to a sample size of 20 (vs. a sample size of 40 in Experiment 1). Thus, the expected frequency for SFSP was 2, while the expected frequency for SFDP, DFSP, and DFDP combined was 18.

Table 3 displays the observed cell frequencies and z values for each

TABLE 3
Frequencies and Fisher z Values for First and Second Choices in Experiment 2

	CELL FREQUENCIES: FIRST CHOICE			CELL FREQUENCIES: SECOND CHOICE		
FACTOR	SPSF	ALL OTHER	z	SFSP	ALL OTHER	z
Extraversion						
High	11	9	3.04**	10	10	2.76**
Low	6	14	1.58	7	13	1.89
Agreeableness						
High	10	10	2.76**	12	8	3.31**
Low	13	7	3.59**	11	7	3.04**
Conscientiousness						
High	10	10	2.76**	2	18	0
Low	5	15	1.25	9	11	2.49*
Emotional Stability						
High	4	16	0.89	5	15	1.25
Low	12	8	3.31**	6	14	1.58
Culture						
High	16	4	4.45**	12	8	3.31**
Low	14	6	3.87**	9	11	2.49*

*$p < .05$.

**$p < .01$.

of the 10 factor–pole combinations for subjects' first and second choices. Significant departures from the expected frequency distribution were seen for 7 of 10 factor–pole combinations for the first choice and 6 of 10 factor–pole combinations for the second choice (all p's $< .05$). Thus, the subjects reported impressions of some kind of correspondence between traits and low or high scores on certain personality tests (factors), despite the fact that no such relationships existed in the stimuli themselves. These impressions were largely consistent with the relationships of those traits to the factors identified by Norman (1963).

EXPERIMENT 3

The first two experiments demonstrated that subjects' intuitions concerning the relationships among personality traits largely conformed to the structure obtained by Norman (1963). This structure appears to be evoked in cognitive tasks performed in a explicitly social context, such as making judgments about particular people (e.g., Norman, 1963), as in the experiments of Passini and Norman (1966), or making judgments about traits in the abstract, as in the studies reported here. Thus, Norman's five factors, and the individual traits that compose them, seem to be organized into operating knowledge structures. Another question concerns the conditions under which these structures may be invoked. Experiment 3 capitalized on the phenomenon of category clustering in verbal learning, in which subjects reorganize an input word list so that the items presented in a random order are recalled in a manner that reflects the pre-existing conceptual relations among them.

METHOD

Subjects

A total of 24 students from introductory psychology classes at the University of Wisconsin–Madison served as subjects, and received extra credit for their participation.

Materials

The 20 positive trait adjectives from Norman (1963) were used to construct lists 20 words in length. A total of 100 lists was prepared, each representing a unique random ordering of the 20 traits, with the

provision that no two traits from the same factor occurred in adjacent positions.

Procedure

Subjects were tested in small groups of two to four subjects in a medium-sized research room divided into two work areas by a cardboard partition. On each of five trials, the experimenter read one of the randomized lists aloud in an even voice at a rate of 3 seconds per word. A different randomized list of the 20 traits was used on each successive trial to prevent subjects from clustering words simply on the basis of rote memory for a particular ordering of words that would occur if the same list were used for each of the five trials. (In addition, each group of subjects heard a unique set of five randomized lists; thus, each randomized list was used for one trial only and for one group of subjects only.) For each trial, the subjects were asked to recall, in writing, as many of the words as possible in any order that the words occurred to them. Subjects were given 2 minutes to perform this task.

After completing the five free-recall trials, each subject was given a set of 20 cards on which the 20 traits were individually typed. Subjects were asked to sort their decks of cards into at least two, but no more than seven, groupings of conceptually related words. An example from the nonsocial domain was provided to illustrate the task. Working individually, the subjects spread their cards out onto a table and sorted the cards into piles. No time limit was imposed. Upon completion of the card-sorting task, the experimenter then asked subjects to examine each grouping and to order the words in each grouping from the most representative word to the least representative word. The task was illustrated by another example from the nonsocial domain. After the subjects finished both the categorization and representativeness judgments, they were debriefed and thanked for participating in the experiment and were given extra-credit points.

RESULTS

The "repetition ratio" (RR), a measure of category clustering in free recall devised by Bousfield (1953), was used to create a dependent measure for the free-recall data. RR is the ratio of the number of repetitions in a subject's recall to the number of words recalled, where a repetition is defined as the adjacent recall of two items from the same category. The advantages of RR over other available measures of

clustering in free recall are that the expected value of RR is easily determined and that the measure is independent of the number of words recalled (Frender & Doubilet, 1974).

For each subject, two RR scores were computed. One RR was computed using the five Norman (1963) personality factors as categories for determining repetitions (this measure is called "consensual RR," or CRR). Another RR (called "idiosyncratic RR," or IRR) was computed using the categories identified by the individual subjects themselves. CRR represented the degree to which the consensual or shared view of personality was used to organize recall of the trait adjectives, whereas IRR estimated the degree to which recall was organized according to the subjects' idiosyncratic views. For some subjects, there would probably be considerable overlap between the two measures, while for other subjects, idiosyncratic views of trait relationships would yield higher clustering as measured by IRR than CRR.

Separate one-sample t tests were performed for CRR and IRR averaged across the five trials to test the hypothesis that the obtained levels of clustering were greater than chance-level clustering (RR = .158). Results of these one-sample t tests showed that the average value of CRR ($M = .164$, $SD = .061$) was not greater than chance, $t\ (23) = 0.48$. However, the value of IRR ($M = .253$, $SD = .134$) was significantly above chance levels, $t\ (23) = 3.47$.

The clustering scores were subjected to a 2×5 repeated-measures analysis of variance (ANOVA) with the two levels of factor type (consensual vs. idiosyncratic categories) and five levels of trials (which corresponded to the five free-recall trials). Both factors were treated as repeated measures. Factor type was treated as a repeated-measures factor because CRR and IRR were both computed from the same subjects' data.

The means and standard deviations for each of the five trials for CRR, IRR, and total number of items recalled are shown in Table 4. There was a significant main effect for category type, $F\ (1, 23) = 13.6$, $p < .01$, such that IRR was consistently higher than CRR. Both the main effect for trials, $F\ (4, 92) = 1.07$, and the category $\times$ trials interaction, $F\ (4, 92) = 1.02$, failed to attain statistical significance. Linear trend analyses on each of the measures revealed a significant increasing trend for total number of items recalled, $F\ (1, 23) = 3.23$, $p < .05$ (one-tailed). Linear trend analyses for the two clustering scores failed to reach statistical significance: for CRR, $F\ (1, 23) = 0.44$; for IRR, $F\ (1, 23) = 2.09$. Thus, there was some degree of organization when clustering was assessed against the idiosyncratic categories, but not when assessed against the consensual "Big Five." Interestingly, the

TABLE 4
Means and Standard Deviations for Recall and Two Measures of Clustering, Experiment 3

	RECALL		CRR		IRR	
TRIAL	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	8.58	1.98	.16	.10	.23	.19
2	10.71	1.92	.19	.11	.23	.16
3	12.50	2.80	.13	.10	.24	.17
4	14.13	2.63	.16	.09	.26	.17
5	15.46	2.69	.18	.10	.28	.19

Note. $n = 24$ for each mean and standard deviation.

level of category clustering for consensual structures was quite low. But recall increased systematically across trials, as expected, while there was no corresponding increase in either consensual or idiosyncratic category clustering. In addition, the amount of consensual clustering observed did not differ from that expected by chance.

EXPERIMENT 4

Although Experiments 1 and 2 demonstrated that subjects' intuitive knowledge of the Norman "Big Five" influenced judgments of trait co-occurrence, Experiment 3 failed to show the influence of these same cognitive structures on the organization of free recall. This unexpected outcome may have been due to the nonsocial nature of the verbal-learning task employed in Experiment 3. A number of investigators (Hamilton, 1981; Hamilton, Katz, & Leirer, 1980; Hoffman, Mischel, & Mazze, 1981; Jeffrey & Mischel, 1979) have reported that the extent of clustering is a function of cognitive task demands. For example, there is more clustering when subjects are asked to form an impression of the target person than when they are simply asked to memorize the list items.

It was hypothesized that embedding the free-recall task in an explicitly social context would facilitate the activation of personality schemas, as reflected in greater consensual category clustering in free recall. To permit better assessment of whatever clustering was observed among the trait terms, a word list comprising easily categorizable nouns unrelated to personality (e.g., "hammer," "toaster," and "magazine") was also used.

METHOD

Subjects

A total of 80 students from introductory psychology classes at the University of Wisconsin–Madison served as subjects, and received extra credit for their participation.

Materials

Two sets of word lists were constructed, using either the positive trait adjectives from Norman (1963) or nouns unrelated to personality that were drawn from categories for which norms were established by Battig and Montague (1969). Four nouns from each of five categories were selected such that the nouns were matched as closely as possible in length to the Norman trait adjectives. Thus, each word list consisted of 20 words categorizable into five categories. The 20 positive trait adjectives are listed in Table 1, and the 20 nouns are listed in Table 5.

For each word type (trait or noun), 100 lists were prepared such that each list consisted of a unique random ordering of the 20 words,

TABLE 5
Noun and Category Relationships in Experiment 4

CATEGORY	NOUNS
Furniture	Table
	Chair
	Television
	Dresser
Reading material	Encyclopedia
	Magazine
	Newspaper
	Paperback
Kitchen utensils	Spatula
	Blender
	Toaster
	Frying pan
Clothing	Shoes
	Raincoat
	Sweater
	Underwear
Tools	Hammer
	Pliers
	Chisel
	Screwdriver

with the provision that no two words from the same category occurred in adjacent positions.

Procedure

Subjects were tested in small groups of two to four students, with seating arranged such that they were unable to see each other's work. The experimenter explained the basic free-recall procedure and, depending on the condition, gave additional instructions concerning the word lists. In the nonsocial conditions (trait or noun), subjects were not told anything about the nature of the words. In the social/trait condition, subjects were told that the words that they would hear were the attributes of a particular person. In the social/noun condition, subjects were told that the words described the possessions of a particular person. Thus, subjects were instructed to form an impression of a person based either on attributes (traits) or possessions (nouns). (As in Experiment 3, a different randomized word list was used for each trial, and each group of subjects heard a unique set of five different randomized word lists.) After the five trials, a detailed debriefing was given by the experimenter, and extra-credit point cards were distributed to the subjects.

RESULTS

The measure of category clustering was CRR, as described in Experiment 3. A 2 (social vs. nonsocial) × 2 (trait vs. noun) × 5 (trials) mixed-design ANOVA was performed on the data. The first two factors were between-subjects factors, while the trials factor was a repeated-measures factor. The means and standard deviations for each condition for each of the five trials are presented in Table 6.

As expected, there was a significant main effect for word type, $F\,(1, 76) = 148.02$, $p < .001$. Given the low level of clustering found in Experiment 3 and the typically high degree of clustering found in noun lists, this result is unremarkable. The main effect for trials was significant, $F\,(4, 304) = 18.86$, $p < .001$, as was the trials × word type interaction, $F\,(4, 304) = 11.98$, $p < .001$. An inspection of the cell means showed that for nouns there was a consistent linear increase across the five trials in both conditions, while the cell means for the traits showed no consistent patterns in either social or nonsocial conditions.

There was also a significant main effect for the social–nonsocial factor, $F\,(1, 76) = 5.57$, $p < .05$, as well as a significant two-way social–nonsocial × word type interaction, $F\,(1, 76) = 4.18$, $p < .05$. An in-

TABLE 6
Means and Standard Deviations for Clustering by Condition and Trial, Experiment 4

		WORD TYPE			
		NORMAN (1963) TRAITS		NOUNS	
SOCIAL–NONSOCIAL	TRIAL	*M*	*SD*	*M*	*SD*
Social	1	.096	.104	.285	.204
	2	.143	.118	.339	.156
	3	.162	.100	.389	.175
	4	.147	.089	.469	.197
	5	.183	.090	.525	.179
Nonsocial	1	.170	.164	.393	.186
	2	.139	.121	.426	.163
	3	.175	.112	.568	.184
	4	.151	.109	.563	.185
	5	.136	.080	.612	.153

Note. $n = 20$ for each cell.

spection of the cell means collapsed across trials revealed that this effect was due primarily to the difference between the two noun conditions. A test on the mean differences between the social ($M = .40$) and nonsocial ($M = .51$) conditions under the noun condition was significant, $F(1, 76) = 9.65$, $p < .01$. The difference between the two trait conditions was negligible. Thus, imposing a social context on a free-recall task involving nouns normally unrelated to personality resulted in a lower level of category clustering.

The remaining interactions failed to attain statistical significance: for the social–nonsocial × trials interactions, $F(4, 304) = 1.84$; for the social–nonsocial × word type × trials interaction, $F(4, 304) = 1.07$.

EXPERIMENT 5

The final experiment tested yet another hypothesis concerning the failure to find consensual category clustering in the free recall of the Norman trait adjectives. It will be recalled that Norman (1963) established his taxonomic structure of personality through factor analyses of peer nominations. The resulting factor structure was described as

"relatively orthogonal and highly stable" (p. 581), but Norman also acknowledged that "occasionally a scale [trait-adjective] was found to load moderately on a factor for which it was not an a priori salient" (p. 581). Intuitively, for example, "poised" would seem to be associated with the category of Culture as well as with the category of Emotional Stability. This overlap in meaning for several of the factors and traits could result in a lowering of category clustering in free recall. In fact, an examination of the intercorrelations among the five factors (obtained by summing factor scores within a factor to create a summative index for each factor for each subject) reported by Norman (1963) reveals that, for some samples, several of the factors were correlated as high as .63 (Conscientiousness with Culture) and .55 (Agreeableness with Emotional Stability) (p. 580). Thus, there appears to be at least a moderate degree of overlap in meaning in the traits in Norman's taxonomy.

METHOD

Subjects

Twenty students from the introductory psychology classes at the University of Wisconsin–Madison served as subjects and received extra credit for their participation.

Materials

A list of all combinations of positive traits with the five personality factors from Norman (1963) was constructed such that no two traits from the same factor were adjacent and no factors were adjacent. The resulting list consisted of 100 trait–factor combinations.

Procedure

Subjects were tested in groups of approximately 10 students who were seated comfortably at large tables in a moderate-sized room. Each subject was provided with a sheet of paper with a 9-point scale at the top of the page and 100 blank lines numbered consecutively from 1 to 100. The end points of the scale were "not at all similar" (1) and "very similar" (9). The experimenter explained that he would be reading the word pairs in the following sentence: "How similar in meaning is *X* to *Y*?", where *X* was replaced by a trait term and *Y* by a factor label. Subjects were asked to record their similarity judgments on the

sheet of paper in the appropriate blank. The experimenter proceeded to read aloud the 100 sentences containing the trait–factor pairs in an even voice at a rate of one sentence every 20 seconds. At the conclusion of the judgmental task, the experimenter collected the sheets, debriefed the subjects, and distributed extra-credit point cards.

RESULTS

Means for each of the 100 possible trait–factor combinations were calculated. The mean similarity rating for each trait on each factor is presented in Table 7. In addition, a mean was calculated for each group of four *a priori* salients. These are also presented in Table 7. For each factor, the within-factor mean (found on the main diagonal) was higher than for the means of the other sets of traits (found off the diagonal). This result suggests that the traits that loaded on a particular factor were more similar in meaning to one another than they were to other traits taken from the other factors. At the same time, however, inspection of the individual trait–factor pair ratings revealed that, for each factor, at least one nonfactor trait had a mean high enough to place it in the top four traits judged most similar to the particular factor. For example, the four traits judged most similar to Agreeableness were "cooperative" ($M=7.650$), "good-natured" ($M=7.050$), "sociable" ($M=6.100$), and "calm" ($M=5.700$). Of these, "sociable" and "calm" were not within-factor traits (the other within-factor traits for Agreeableness were "not jealous," $M=4.300$, and "gentle," $M=4.650$). Also, both "sociable" and "calm" had high similarity on their respective factors as well. Thus, it appeared that certain traits within the Norman (1963) list overlapped in meaning with two of the personality factors.

RECONCILING DISCREPANT RESULTS

The results of Experiments 1 and 2 strongly suggested that subjects used the Norman factors as schemas in making both *a priori* and memory-based judgments of trait co-occurrence. Nevertheless, Experiments 3 and 4 equally strongly indicated that subjects failed to use the same five factors as schemas for organizing recall of a list of trait adjectives. Experiment 5 showed that there was some lack of semantic distinctiveness or factor coherence among the five Norman factors. Accordingly, the results of the first four experiments were reanalyzed to determine whether this lack of distinctiveness might account for

the failure of the five factors to act consistently like schemas in organizing recall.

PREDICTING CO-OCCURRENCE RATINGS

Experiments 1 and 2 showed that two traits drawn from the same pole of a particular factor were judged more likely to co-occur than were traits drawn from opposite poles, or from the other four factors. However, the aggregate scores used in this analysis may have obscured differences across traits in judged co-occurrence. That is, only a few perceived co-occurrences for each factor may have accounted for the overall finding in Experiments 1 and 2 that subjects appeared to use the Norman structure. As an initial approach to this problem, the trait–factor similarity judgments collected in Experiment 5 were correlated with the trait–factor co-occurrence judgments collected in Experiments 1 and 2. A high correlation would indicate that subjects in Experiments 1 and 2 tended to select traits that were judged in Experiment 5 to be highly similar to a corresponding "Big Five" factor.

In order to provide a common metric, the 100 (20 positive traits × 5 factors) trait–factor similarity ratings collected in Experiment 5 were converted into z scores. These standard scores were based on within-factor means and standard deviations in order to standardize the ratings within factors rather than across factors. Thus, within each of the five factors, positive z scores would correspond to high similarity ratings, and negative z scores would correspond to low similarity ratings. In addition, the co-occurrence ratings collected in Experiments 1 and 2 were used to create an index representing the percentage of times a given trait was chosen as an instance of its corresponding "Big Five" factor, and the appropriate pole of that factor, relative to the total number of times the trait was chosen overall. This index provided a measure of the degree to which subjects judged particular traits as corresponding to the factor–pole combination predicted by the Norman "Big Five" structure. (For example, in Experiment 1, "talkative" was chosen 13 times overall and was judged to co-occur with the positive pole of the Extraversion factor 11 times. The index value would be $11/13 = 0.85$. Put another way, 85% of the subjects who chose "talkative" for any of the factors paired "talkative" with the positive pole of the Extraversion factor.)

The correlation between the relative similarity index of Experiment 5 and the relative co-occurrence index of Experiment 1 was highly significant, $r\,(38) = .68$, $p < .001$. The corresponding correlation for Experiment 2 was in the same direction, although it did not reach

TABLE 7
Means for Similarity Judgments for the Norman (1963) Traits and Factors, Experiment 5

	FACTORS				
TRAIT SETS	EXTRAVERSION	AGREEABLENESS	CONSCIENTIOUSNESS	EMOTIONAL STABILITY	CULTURE
Extraversion					
Talkative	6.750*	3.050	2.650	3.600	3.100
Frank	4.500	2.950	3.850	3.100	2.750
Adventurous	7.000*	2.650	2.250	3.900	3.400
Sociable	6.000*	6.100*	4.200	5.100	4.650
(Mean)	(6.063)	(3.688)	(3.238)	(3.925)	(3.475)
Agreeableness					
Good-natured	4.650	7.050*	4.950	4.850	3.450
Not jealous	4.000	4.300	3.050	6.250*	2.400
Gentle	2.450	4.650	4.350	4.450	3.600
Cooperative	5.000	7.650*	5.300*	4.450	3.750
(Mean)	(4.025)	(5.913)	(4.413)	(5.000)	(3.300)

Conscientiousness					
Tidy	2.950	3.400	6.250*	2.900	3.050
Responsible	2.750	3.950	7.400*	5.600	4.250
Scrupulous	3.950	2.650	4.800	3.050	2.650
Persevering	4.150	3.150	4.850	4.350	3.450
(Mean)	(3.450)	(3.288)	(5.825)	(3.975)	(3.350)
Emotional Stability					
Poised	3.300	4.100	5.100	6.800*	5.300*
Calm	2.250	5.700*	4.550	7.500*	3.900
Composed	2.900	4.800	5.150*	7.000*	4.350
Not hypochondriacal	3.050	3.800	3.200	5.050	3.250
(Mean)	(2.875)	(4.600)	(4.500)	(6.588)	(4.200)
Culture					
Artistically sensitive	3.050	3.000	3.050	3.650	5.200*
Intellectual	3.350	3.000	4.450	3.800	4.950*
Refined	3.000	4.150	5.000	4.950	5.700*
Imaginative	5.150*	2.800	3.100	3.150	3.500
(Mean)	(3.133)	(3.238)	(3.900)	(3.888)	(4.838)

Note. Asterisks indicate the four highest trait means for a factor.

conventional levels of statistical significance, r (18) = .39. Thus, the subjects' co-occurrence judgments were not distributed evenly across all traits on the same pole of each factor. Rather, they tended to be confined to a few traits within each factor—those judged to be more similar in meaning to the factor name.

CLUSTERING OF CO-OCCURRENCE CHOICES

In Experiments 1 and 2, subjects chose two traits as co-occurring with each factor. In the original analysis, these choices were analyzed separately, without regard to any relations between them. If these co-occurrence judgments were influenced by the five-factor structure, it would be expected that both choices would come from the same pole of the factor in question.

In order to address this question, subjects' first and second choices of traits were categorized as to whether they both came from the same pole of the same factor. Because subjects had the entire list of 40 trait adjectives to examine when making their co-occurrence judgments, one would expect that the two choices would strongly tend to be drawn from the same pole of the same factor. Table 8 presents the results of this analysis for both Experiments 1 and 2. For Experiment 1, the mean percentage of subjects whose first and second choices both came from the positive pole of a factor was 29.60 (SD = 9.79), while the mean percentage of subjects whose first and second choices both came from the negative pole of a factor was 41.60 (SD = 7.77). For Experiment 2, the mean percentages were 22.00 (SD = 18.57) and 23.00

TABLE 8
Proportion of Subjects Making SFSP First and Second Choices in Experiment 1 and Experiment 2

	EXPERIMENT 1		EXPERIMENT 2	
FACTOR	POSITIVE TRAITS	NEGATIVE TRAITS	POSITIVE TRAITS	NEGATIVE TRAITS
Extraversion	16/40 (40%)	22/40 (55%)	5/20 (25%)	2/20 (10%)
Agreeableness	13/40 (33%)	16/40 (40%)	5/20 (25%)	7/20 (35%)
Conscientiousness	6/40 (15%)	16/40 (40%)	1/20 (5%)	3/20 (15%)
Emotional Stability	10/40 (25%)	14/40 (35%)	1/20 (5%)	4/20 (20%)
Culture	14/40 (35%)	15/40 (38%)	10/20 (50%)	7/20 (35%)

($SD = 11.51$), respectively. While there was somewhat more clustering of negative than positive traits, the overall levels of clustering appeared to be relatively low. This indicates that while certain individual traits may have been strongly linked to particular factors, their links to other traits were substantially weaker; in other words, the traits themselves were not strongly linked to each other.

ITEM-BY-ITEM CLUSTERING ANALYSIS

The previous reanalyses indicated that the five factors were far from homogeneous with respect to semantic similarity and judged co-occurrence. Of all the traits located on the same pole of a factor, certain ones were far more likely to be associated with that factor than others. Such internal inconsistencies may have resulted in low overall levels of clustering in Experiments 3 and 4. Accordingly, it seemed important to determine precisely which traits were clustered together in Experiments 1 and 2 and in Experiments 3 and 4. For the purposes of this analysis, data for Experiments 1 and 2 were combined, as were data for Experiments 3 and 4. Only subjects who studied trait names in the nonsocial context of Experiment 4 were combined with data in Experiment 3, in order to preserve comparability across experiments. Table 9 presents the frequency of all within-factor pairings for the positive traits in the "Big Five" structure.

The left-hand column shows the clustering data from Experiments 1 and 2, analyzed in the preceding section. It is apparent that only a few trait pairs account for the vast majority of clustering observed in Experiments 1 and 2. For example, "talkative" and "sociable" accounted for 10 of the 21 within-factor pairings on Extraversion. Similarly, "cooperative" and "good-natured" accounted for 15 of 18 clusters on Agreeableness, while "artistically sensitive" and "intellectual" accounted for 14 of 24 clusters on Culture. There were no prominent clusters observed on the factors of Conscientiousness or Emotional Stability. Somewhat different results were obtained in Experiments 3 and 4 (shown in the right-hand column of Table 9). As in Experiments 1 and 2, "sociable" and "talkative" accounted for most of the clustering observed on the Extraversion factor. However, the few clusters observed in Experiments 3 and 4 were distributed more evenly across the relevant traits.

Thus, the subjects tended to cluster trait pairs in different ways, depending on the experimental task. Still, reanalysis of clustering in Experiments 1 and 2 suggested that there may be weak links between

TABLE 9
Clustering Pair Frequencies for Experiments 1–2 and Experiments 3–4

FACTOR	EXPERIMENTS 1–2[a]	EXPERIMENTS 3–4[b]
Extraversion		
Sociable–talkative	10	15
Sociable–adventurous	6	2
Talkative–adventurous	4	6
Frank–adventurous	0	3
Talkative–frank	0	1
Sociable–frank	1	1
Agreeableness		
Cooperative–good-natured	15	2
Cooperative–not jealous	1	3
Cooperative–gentle	1	2
Good-natured–gentle	1	7
Good-natured–not jealous	0	4
Not jealous–gentle	0	1
Conscientiousness		
Responsible–scrupulous	4	3
Responsible–persevering	2	1
Responsible–tidy	1	2
Scrupulous–tidy	0	5
Scrupulous–persevering	0	5
Persevering–tidy	0	2
Emotional Stability		
Composed–calm	5	5
Composed–poised	3	7
Calm–poised	2	2
Not hypochondriacal–calm	1	2
Not hypochondriacal–poised	0	4
Not hypochondriacal–composed	0	1
Culture		
Artistically sensitive–intellectual	14	3
Artistically sensitive–refined	5	2
Artistically sensitive–imaginative	3	2
Refined–intellectual	2	1
Imaginative–intellectual	0	8
Refined–imaginative	0	1

[a] $n = 60$.

[b] Data from subjects in Experiment 3 ($n = 24$) and the corresponding condition in Experiment 4 (nonsocial/trait condition; $n = 20$) were combined. Total pairings for these 44 subjects = 682.

the subordinate traits for the Norman factors. Apparently, even traits drawn from the same pole of a particular factor are not considered highly similar in meaning. If, in fact, subordinate traits in the "Big Five" are not as interchangeable as previously thought, the failure to find above-chance-level clustering in the free-recall experiments is no longer surprising.

DISCUSSION

The present experiments tested the hypothesis that the five major personality factors described by Norman (1963) and others constitute schemas, or cognitive structures, that are available to guide the processing of social information. Some of the present experiments supported the hypothesis that the "Big Five" factors operate as cognitive schemas. In Experiment 1, which was based on the conceptual-similarity methodology of Shweder and D'Andrade (e.g., 1979), subjects' memory-based judgments concerning the relations between primary traits and superordinate factors almost perfectly paralleled the structure found in Norman's data. Experiment 2, which was based on the illusory-correlation phenomenon, showed that subjects' memory-based judgments of co-occurrence between traits and factors also reflected, to a considerable degree, the trait–factor relations predicted by the "Big Five." Because of the very large number of items presented in Experiment 2, the subjects were almost certainly prevented from relying on rote memory for the stimuli, and therefore were forced to make inferences based on their intuitive theories concerning personality–trait relationships—theories revealed by Experiment 1 to match the "Big Five." Also, reanalyses showed that the patterns of co-occurrence judgments in Experiments 1 and 2 were predictable to some extent from similarity ratings obtained in Experiment 5.

In other respects, however, the "Big Five" failed to guide social information processing. In Experiment 3, in which the category-clustering methodology of Bousfield (1953) was used, there was no evidence that subjects utilized the "Big Five" as schemas for organizing recall of trait terms in a verbal-learning task. Experiment 4 confirmed this result, even though the verbal-learning procedure was embedded in an impression-formation task that should have made the personality connotations of the list items more salient, thus increasing the probability that the "Big Five" schemas would be activated. In brief, the "Big Five" structure appeared to act like a powerful cognitive structure in tasks such as conceptual-similarity ratings (Experiment 1) and co-occurrence judgments (Experiment 2), but was apparently inactive when it failed to organize free recall of personality traits (Experiments 3 and 4).

A reanalysis of Experiments 1 and 2, in which a new measure of clustering was computed, showed that the level of category clustering among ostensibly related trait terms was much lower than expected, especially for the positive traits. In these experiments, subjects would be expected to show high levels of clustering (making many SFSP choices) if the Norman factors were operating as schemas. It should

be noted that subjects had the entire list of traits available during their judgments in these experiments, in contrast to Experiments 3 and 4, where subjects were required to recall traits from memory. But subjects showed fairly low levels of category clustering even under the nearly ideal conditions in Experiments 1 and 2. This finding may explain the low levels of clustering found in Experiments 3 and 4.

While the results of Experiments 1 and 2 are consistent with a wealth of research showing that the multidimensional structures derived from memory-based ratings are largely replicable in conceptual similarity ratings (Shweder, 1982; Shweder & D'Andrade, 1979, 1980), the low levels of category clustering in Experiments 3 and 4 are consistent with other research employing clustering methodology. Hamilton (Hamilton, 1981; Hamilton *et al.*, 1980), Mischel (Hoffman *et al.*, 1981; Jeffrey & Mischel, 1979), and their colleagues have reported that organizational activity of the sort represented by category clustering varies across conditions. However, these investigators have consistently failed to find high levels of clustering based on social stereotype categories (see also Hastie, Park, & Weber, 1984). For example, the maximum clustering obtained in a study by Hamilton, Leirer, and Katz (1979, cited in Hamilton, 1981, pp. 143–144) reached only about 1.75 using the Bousfield and Bousfield (1966) measure, for a mean of 11 items recalled from a 16-item list (four categories with four exemplars each) representing categories of personality content. The theoretical maximum clustering value for a mean recall of 11 items, assuming each of the four categories was represented about equally, would be approximately 5.25[1]. Thus, Hamilton's highest clustering value fell far short of the high clustering values typically obtained with nonsocial items.

It is possible that the "Big Five" are simply too loosely organized to function reliably as cognitive schemas. In Experiments 3 and 4, the levels of category clustering observed for trait terms were substantially

1. The Bousfield and Bousfield (1966) clustering measure (*BBD*) is calculated as the number of repetitions (R) in a recall list minus the expected value of the repetitions [$E(R)$]. Thus, $BBD = R - E(R)$. The $E(R)$ is calculated according to the following formula: $E(R) = (n/N) - 1$, where n is the number of items from each category that are recalled and N is the total number of items recalled. Given Hamilton's stimulus set of 16 items with four categories, each having four exemplars, a recall list with maximum clustering (assuming about equal representation from each category and recall of 11 items) might look like this: aaabbbcccdd, where each letter (e.g., "a") represents the sequential occurrence of a category exemplar. In this theoretical list, there are seven repetitions ($R = 7$), and the expected value of R is about 1.75. Thus, given maximum clustering in an 11-item recall list, the clustering value would be as follows: $BBD = 7 - 1.75 = 5.25$.

lower than those observed with the nonsocial terms. In addition, the levels of clustering observed for the nouns in Experiment 4 were typical of those found in verbal-learning experiments (Bousfield, 1953; Wilson & Kihlstrom, 1986). The categories used in verbal-learning experiments involving nonsocial terms typically comprise highly representative category instances, and there is usually a high degree of consensus among subjects as to category membership. This means that subjects are very unlikely to assign an item to a category that is different from the one intended by the experimenter.

In contrast to the usual case with nouns, Experiment 5 revealed a considerable lack of distinctiveness in the relationships of instances to categories within the personality domain, implying that the category assignments of subjects frequently differed from the category assignments predicted by factor analysis. Even though the subordinate traits selected by Norman were those that had the highest loadings on their respective superordinate factors, the semantic associations between subordinate and superordinate trait concepts may be very weak. If so, representations of the superordinate categories would not be reliably activated by presentation of subordinate items. In Experiment 5, the trait–factor relations were largely as expected for the Extraversion, Emotional Stability, and Culture factors (though even in these cases, one trait on each factor failed to line up as expected). However, the remaining two factors—Agreeableness and Conscientiousness—failed to mirror the structure retrieved by factor analysis.

Reanalysis of Experiments 1 and 2 shed further light on this problem by revealing that the close matches between the "Big Five" factor structure on the one hand, and the conceptual and co-occurrence structures on the other, were somewhat illusory. Thus, most of the critical SFSP responses given in those experiments were accounted for primarily by three factors, and by only a few traits within those factors. The answer to the question posed in the title of this paper—"When is a schema not a schema?"—appears to be that *some* of the "Big Five" factors operate reliably as cognitive schemas *some* of the time. Not all the trait–factor relations revealed by factor analysis appear to be represented in people's implicit personality theories. And even the strongest conceptual relations appear to be rather weak—far weaker than those observed in the nonsocial domain of natural objects. So only some traits will reliably activate a "Big Five" schema, and this activation will not necessarily include all the traits identified as most representative of the factor structure. When other traits are presented, subjects will be on their own, relying on implicit theories of personality quite different from the "Big Five."

REFERENCES

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

Anderson, J. R. (1985). *Cognitive psychology and its implications* (2nd ed.). San Francisco: W. H. Freeman.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs, 80*(3, Pt. 2).

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, England: Cambridge University Press.

Block, J., Weiss, D. S., & Thorne, A. (1979). How relevant is a semantic similarity interpretation of personality ratings? *Journal of Personality and Social Psychology, 37*, 1055–1074.

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology, 49*, 229–240.

Bousfield, A. K., & Bousfield, W. A. (1966). Measurement of clustering and sequential constancies in repeated free recall. *Psychological Reports, 19*, 935–942.

Brewer, W. F., & Nakamura, G. V. (1984). The nature and functions of schemas. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 119–160). Hillsdale, NJ: Erlbaum.

Bruner, J. S., & Tagiuri, R. (1954). Person perception. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 2, pp. 634–654). Reading, MA: Addison-Wesley.

Buss, D. M., & Craik, K. H. (1985). Why not measure that trait? Alternative criteria for identifying important criteria. *Journal of Personality and Social Psychology, 48*, 934–946.

Cantor, N., & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. *Journal of Psychology and Social Psychology, 35*, 38–48.

Cantor, N., & Mischel, W. (1979a). Prototypes in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 12, pp. 3–52). New York: Academic Press.

Cantor, N., & Mischel, W. (1979b). Prototypicality and personality: Effects on free recall and personality impressions. *Journal of Research in Personality, 13*, 187–205.

Cattell, R. B. (1947). Confirmation and clarification of primary personality factors. *Psychometrika, 12*, 197–220.

Cattell, R. B. (1957). *Personality and motivation structure and measurement*. Yonkers-on-Hudson, NY: World Book.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 74*, 271–280.

Digman, J. M., & Inouye, J. (1986). Further specification of the five robust factors of personality. *Journal of Personality and Social Psychology, 50*, 116–123.

Digman, J. M., & Takemoto-Chock, N. K. (1981). Factors in the natural language of personality: Re-analysis and comparison of six major studies. *Multivariate Behavioral Research, 16*, 149–170.

Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach* (Perspectives on Individual Differences Series). New York: Plenum Press.

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *Journal of Abnormal and Social Psychology, 44*, 329–344.

Frender, R., & Doubilet, P. (1974). More on measures of category clustering in free recall—although not the last word. *Psychological Bulletin, 81*, 64–66.

Goldberg, L. R. (1981). From ace to zombie: Some explorations in the language of personality. In C. Spielberger & J. N. Butcher (Eds.), *Advances in personality assessment* (Vol. 1, pp. 203–234). Hillsdale, NJ: Erlbaum.

Hakel, M. D. (1969). Significance of implicit personality theories for personality research and theory [Summary]. *Proceedings of the 77th Annual Convention of the American Psychological Association*, 403–404.

Hamilton, D. L. (1981). Cognitive representations of persons. In E. T. Higgins, P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario Symposium* (Vol. 1, pp. 135–159). Hillsdale, NJ: Erlbaum.

Hamilton, D. L., Katz, L. B., & Leirer, V. O. (1980). Organizational processes in impression formation. In R. Hastie, T. M. Ostrum, E. B. Ebbesen, R. S. Wyer, D. L. Hamilton, & D. E. Carlston (Eds.), *Person memory: The cognitive basis of social perception* (pp. 121–152). Hillsdale, NJ: Erlbaum.

Hampson, S. E., John, O. P., & Goldberg, L. R. (1986). Category breadth and hierarchical structure in personality: Studies of asymmetries in judgments of trait implication. *Journal of Personality and Social Psychology, 51*, 37–54.

Hastie, R. (1980). Memory for behavioral information that confirms or contradicts a personality impression. In R. Hastie, T. M. Ostrom, E. B. Ebbesen, R. S. Wyer, D. L. Hamilton, & D. E. Carlston (Eds.), *Person memory: The cognitive basis of social perception* (pp. 155–177). Hillsdale, NJ: Erlbaum.

Hastie, R., Park, B., & Weber, R. (1984). Social memory. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 2, pp. 151–212). Hillsdale, NJ: Erlbaum.

Hoffman, C., Mischel, W., & Mazze, K. (1981). The role of purpose in the organization of information about behavior: Trait-based versus goal-based categories in person cognition. *Journal of Personality and Social Psychology, 40*, 211–225.

Hogan, R. T. (1983). Socioanalytic theory of personality. In M. M. Page (Ed.), *Nebraska Symposium on Motivation, 1982: Personality—Current theory and research* (Vol. 30, pp. 55–89). Lincoln: University of Nebraska Press.

Jackson, D. N. (1982). Some preconditions for valid person perception. In M. P. Zanna, E. T. Higgins, & C. P. Herman (Eds.), *Consistency in social behavior: The Ontario Symposium* (Vol. 2, pp. 251–279). Hillsdale, NJ: Erlbaum.

Jeffery, K., & Mischel, W. (1979). Effects of purpose on the organization and recall of information in person perception. *Journal of Personality, 47*, 397–419.

Kim, M. P., & Rosenberg, S. (1980). Comparison of two structural models of implicit personality theory. *Journal of Personality and Social Psychology, 38*, 375–389.

Kripke, S. (1972). Naming and necessity. In D. Davidson & G. Harman (Eds.), *Semantics of natural language* (pp. 253–355). Boston: Reidell.

Marascuilo, L. A., & McSweeney, M. (1977). *Nonparametric and distribution-free methods for the social sciences*. Monterey, CA: Brooks/Cole.

McCrae, R. R., & Costa, P. T. (1985a). Comparison of EPI and psychoticism scales with measures of the five-factor model of personality. *Personality and Individual Differences, 6*, 587–597.

McCrae, R. R., & Costa, P. T. (1985b). Updating Norman's "adequate taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of Personality and Social Psychology, 49*, 710–721.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. San Francisco: W. H. Freeman.

Newcomb, T. (1931). An experiment designed to test the validity of a rating technique. *Journal of Educational Psychology, 22,* 279–289.

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.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning.* Urbana: University of Illinois Press.

Passini, F. T., & Norman, W. T. (1966). A universal conception of personality structure? *Journal of Personality and Social Psychology, 4,* 44–49.

Putnam, H. (1975). The meaning of meaning. In H. Putnam, *Mind, language, and reality: Philosophical papers* (Vol. 2, pp. 215–271). Cambridge, England: Cambridge University Press.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 28–48). Hillsdale, NJ: Erlbaum.

Rosenberg, S. (1977). New approaches to the analysis of personal constructs in person perception. In A. W. Landfield (Ed.), *Nebraska Symposium on Motivation, 1976* (Vol. 24, pp. 179–242). Lincoln: University of Nebraska Press.

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology, 9,* 283–294.

Rosenberg, S., & Sedlak, A. (1972). Structural representations of implicit personality theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 235–297). New York: Academic Press.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence, and education* (pp. 33–58). Hillsdale, NJ: Erlbaum.

Schneider, D. J. (1973). Implicit personality theory: A review. *Psychological Bulletin, 79,* 294–309.

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. (1979). Accurate reflection or systematic distortion? A reply to Block, Weiss, and Thorne. *Journal of Personality and Social Psychology, 37,* 1075–1084.

Shweder, R. A., & D'Andrade, R. G. (1980). The systematic distortion hypothesis. In R. A. Shweder (Ed.), *Fallible judgment in behavioral research* (New Directions for Methodology of Social and Behavioral Science, No. 4, pp. 37–58). San Francisco: Jossey-Bass.

Taylor, S. E., & Crocker, J. C. (1981). Schematic bases of social information processing. In E. T. Higgins, P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario Symposium* (Vol. 1, pp. 89–133). Hillsdale, NJ: Erlbaum.

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology, 4,* 25–29.

Tupes, E. C., & Christal, R. E. (1961). *Recurrent personality factors based on trait ratings* (USAF ASD Technical Report No. 61–97).

Weiss, D. S., & Mendelsohn, G. A. (1986). An empirical demonstration of the implausibility of the semantic similarity explanation of how trait ratings are made and what they mean. *Journal of Personality and Social Psychology, 50,* 595–601.

Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Reading, MA: Addison-Wesley.

Wilson, L., & Kihlstrom, J. F. (1986). Subjective and categorical organization of recall during posthypnotic amnesia. *Journal of Abnormal Psychology, 95*, 264–273.