12-1986

The Linguistic Relativity of Person Cognition: An English-Chinese Comparison

Curt HOFFMAN
University of Alberta

Ivy Yee-Man LAU
Singapore Management University, ivylau@smu.edu.sg

David R. JOHNSON
University of Alberta

DOI: https://doi.org/10.1037/0022-3514.51.6.1097

The linguistic relativity of person cognition: An English-Chinese comparison

Published in Journal of Personality and Social Psychology, 1986 December, Vol. 51, Issue 6, pp. pp. 1097-1105.

https://doi.org/10.1037/0022-3514.51.6.1097

By: Curt Hoffman, *University of Alberta, Edmonton, Canada*;
Ivy Lau, *University of Alberta, Edmonton, Canada*
David Randy Johnson, *University of Alberta, Edmonton, Canada*

Acknowledgement: We thank Igor Gavanski, Yvonne Ko, Judy Lau, Maria Tchir, and Mary Yu for their assistance with the research reported in this article.

The idea that the particular language one speaks importantly affects the manner in which one perceives and thinks about the world—the linguistic relativity hypothesis—has a long but somewhat checkered history within the disciplines of psychology, anthropology, linguistics, and philosophy. Benjamin Lee Whorf, this century's most influential proponent of the linguistic relativity hypothesis, expressed its central proposition as follows:

We dissect nature along lines laid down by our native languages. The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face; on the contrary, the world is presented in a kaleidoscopic flux of impressions which has to be organized by our minds—and this means largely by the linguistic systems in our minds.… We are thus introduced to a new principle of relativity, which holds that all observers are not led by the same physical evidence to the same picture of the universe, unless their linguistic backgrounds are similar, or can in some way be calibrated. (Whorf, 1956, pp. 213–214)

Whorf's writings were the primary inspiration for a small flurry of behavioral research on the language-cognition relation that arose in the 1950s and continued into the 1960s, most of which was concerned with the impact of language on color memory and object classification. By the 1970s, the linguistic relativity hypothesis had largely fallen into disfavor. This was due in part to the vague and sometimes extreme form in which Whorf had stated the hypothesis. It also occurred because anthropology had begun to accord more importance to linguistic and cultural universals than to linguistic and cultural differences (e.g., Berlin & Kay, 1969), and psychology had begun to question (both on empirical and on conceptual grounds) the assumption that categorization is arbitrary and was moving toward the view that the perceptual world contains its own builtin category structure (e.g., Rosch, 1974). Recently, however, there have been signs of renewed interest in the linguistic relativity hypothesis. More sophisticated theoretical analyses of the language-cognition relation (e.g., Bloom, 1981), as well as methodologically superior empirical work (e.g., Lucy & Shweder, 1979), are helping to clarify the ways in which linguistic classifications and linguistic structures can and do partially shape the cognitive categories and processes by which we come to know the world.

This revival of interest in the linguistic relativity hypothesis is particularly timely in view of the central conceptual role currently accorded to the schema concept in cognitive, social, and developmental psychology. As Bloom (1981) has argued, language affects cognitive life in two general ways. First, it influences the development of our repertory of cognitive schemas:

In addition … to developing a large number of schemas free of the influence of language, which never come to be labeled, and to developing a large number of schemas free of the influence of language, which come in time to be labeled, but whose internal organization remain unaffected by the fact that they receive

labels, the child will construct or reconstruct a very large number of schemas expressly to meet the requirements of linguistic labels. (Bloom, 1981, p. 66)

An example of this last type of schema is that labeled by the word *dog;* dogs do not constitute an especially coherent or distinctive category, and were it not for the necessity of learning the correct use of the word *dog*, it is unlikely that the child would ever arrive at this particular equivalence class. Other, more abstract examples include the schemas labeled by the words *sister, amount,* and *tao*.

Second, those schemas that have linguistic labels enjoy a special status in our mental life. According to Bloom, this is because the use of verbal symbols not only is a prerequisite to overt social communication but also facilitates the covert self-communication in which we engage to structure our own thought processes.

We seem to call specially upon those of our schemas that have names, via their names, when we want to disengage particular schematic perspectives from the collectivity of our interacting associations, ideas, and experiences and make use of those discrete, structured perspectives on reality as stable points of mental orientation to provide direction to our continuing cognitive activities. (Bloom, 1981, p. 76)

There is now a good deal of evidence attesting to the special role played by labeled schemas in such diverse cognitive activities as object classification (Carroll & Casagrande, 1958; Greenfield, Reich, & Olver, 1966), memory for forms (Carmichael, Hogan, & Walter, 1932), creative problem solving (Higgins & Chaires, 1980), and deductive reasoning (Clark, 1969), among others (see Bloom, 1981, for a review). Perhaps the best-known research on the power of linguistic labels to guide mental activity is the series of studies on color memory initiated by Brown and Lenneberg (1954). Collectively, this research has demonstrated that those colors that are most readily and most accurately codable (i.e., those that have accessible, accurate verbal labels) are also the colors that people are best able to remember when tested for their recognition memory (e.g., Brown & Lenneberg, 1954; Lantz & Stefflre, 1964; Lucy & Shweder, 1979; Stefflre, Castillo Vales, & Morley, 1966).

Furthermore, given that language affects the acquisition and use of our cognitive schemas, it follows that different languages, because they label certain perspectives on the world but not others, must affect their speakers' repertories of schemas in language-specific ways. This, of course, is simply the linguistic relativity hypothesis restated in contemporary terms.

Surprisingly, the linguistic relativity hypothesis has never, to our knowledge, been tested in the domain of social cognition. Indeed, it is ironic that researchers in the linguistic relativity tradition have typically sought to demonstrate effects of language on thought in domains (such as color memory) where language would seem intuitively to play only a minor role. If linguistic effects on cognitive activity can be demonstrated even here, however, then the effects of language on the more subjective, less perceptually grounded kinds of thinking involved in our transactions with the social world must almost certainly be still more pronounced. This study was therefore designed to test whether distinct languages (in this case, English and Chinese) are capable of exerting language-specific effects on people's impressions of and memory for other individuals.

One important way in which the social lexicons of languages vary is their codification of individual differences, that is, their repertories of labeled schemas for personality traits and types. No language has names for all possible behavior patterns, and the particular personality traits and types singled out for labeling differ somewhat from language to language. For example, what is the English term for the type of personality characterized by these attributes: worldly, experienced, socially skillful, devoted to his or her family, and somewhat reserved? Most readers will agree that there is no economical expression in English that names this particular constellation of attributes. On the other hand, it is just this constellation

of attributes that is named by the Chinese term *shì gù*. Thus, Chinese possesses a relatively short, precise label summarizing a cluster of personality qualities that English can express only via a long string of adjectives. Does this linguistic difference imply that Chinese speakers have readier access to a cognitive schema corresponding to that personality type and would, when confronted with an individual displaying that configuration of attributes, be more likely to demonstrate the sorts of cognitive responses associated with schematic processing?

To address this question, we first identified two personality schemas that, like the one named by *shì gù*, have economical labels in Chinese but not in English, and two that have economical labels in English but not in Chinese. We then constructed parallel English- and Chinese-language descriptions of characters exemplifying these four personality types. Pretesting established that each character's personality was, in fact, more codable (i.e., more readily and accurately nameable) in one language than in the other. The following three groups of subjects were then presented with the descriptions and were later tested for their impressions of and memory for the characters: English monolinguals, Chinese-English bilinguals who read and responded in English, and Chinese-English bilinguals who read and responded in Chinese. We predicted that subjects working in English would show greater evidence of schematic thinking in the case of the two characters representing schemas labeled in English but not in Chinese, and that subjects working in Chinese would show greater evidence of schematic thinking in the case of the two characters representing schemas labeled in Chinese but not in English.

We looked for the following specific indicators of schematic processing in our study (see Fiske & Taylor, 1984). First, schematic processing typically involves "going beyond the information given" and, in the case of person cognition, leads the perceiver to infer schema-congruent attributes that were never directly observed. A second, related consequence of schematic processing is that it often leads the perceiver to misremember schema-congruent but never-seen attributes or behaviors as in fact having been displayed by the person and, more broadly speaking, often impairs the perceiver's later ability to distinguish presented and nonpresented information (e.g., in tests of recognition memory). Third, schematic processing has occasionally been demonstrated to facilitate free recall of information about the person (or at least the ability to reconstruct such information), particularly when the person is a highly prototypical example of the schema (e.g., Cantor & Mischel, 1979b; Forgas, 1983).

A note regarding our choice of subject groups is in order at this point. Most studies in the linguistic relativity tradition have compared monolingual, or nearly monolingual, speakers of different languages. A problem with this type of study is that one cannot separate any effects of language per se from effects due to nonlinguistic cultural differences between the groups. The vast majority of English and Chinese monolinguals, for example, differ in many ways in addition to the language they speak. Our solution to this problem was to assign Chinese-English bilinguals randomly to think and respond in one or the other language during the experiment. A group of English monolinguals from the same university population was also studied. (Unfortunately, a sample of Chinese monolinguals was not available to us, let alone a sample demographically similar to the other groups. The inclusion of such a group was not, however, essential to the logic of the design.) The advantages of this design are that it allows true experimental control over the linguistic variable (at least in the case of the bilinguals) and enables one to isolate linguistic effects from more general cultural effects. In other words, to the extent that any obtained results are due to language per se, the English monolinguals and the Chinese-English bilinguals working in English should respond similarly and should differ from the Chinese-English bilinguals working in Chinese. A hidden assumption underlying this type of design is that bilinguals possess two relatively separate language codes, and that when utilizing one code, there is no strong, automatic tendency to access the other code and its associated schematic knowledge. If, on the other hand, bilinguals do tend to access both codes more or less simultaneously, then of course no effects of language would be obtained with a design such as this.

Pretest Studies: Development of the Stimulus Character Descriptions

Method

The behavioral descriptions used as stimuli in the main experiment were developed through a series of pretest studies. In the first of these, 18 English monolinguals were presented with 16 English-language personality trait terms and were asked to list as many attributes as they could think of, such as typical behaviors, attitudes, abilities, likes and dislikes, and so forth, that are associated with each of these personality types (cf. Cantor & Mischel, 1979a). Similarly, 18 Chinese-English bilinguals listed (in Chinese) attributes associated with the personality types named by 16 Chinese-language trait terms. On the basis of these data and our own intuitions concerning the connotative meanings of the terms, several personality types were selected for further study: two types named by English terms and two named by Chinese terms that appeared to have no corresponding label in the other language; two types that, although labeled in both English and Chinese, seemed to be missing an attribute in one language relative to the other; and two types that are labeled in both English and Chinese and appeared to have essentially identical attributional definitions in each language. After identifying clusters of related attributes associated with each of these eight personality types, we selected the three largest or most important clusters for each type and wrote 6 statements (each describing a specific behavior, attitude, ability, or other concrete attribute) for each cluster. In this way, character descriptions exemplifying the eight personality types were created, each of which consisted of 18 statements. All of the characters were young men identified only by an English first name.

The original drafts of these descriptions were written primarily in English, and great care was taken to use the most concrete language possible and to avoid the use of trait terms or other expressions that do not have more or less exact equivalents in Chinese. Next, each description was translated into Chinese. Changes to the English originals were made whenever necessary to achieve equivalence of meaning. As a final step, the Chinese translations were back-translated into English by bilinguals unconnected with our research, and these back-translated descriptions were compared with the original English versions (cf. Brislin, 1970). This step resulted in a number of further changes to the materials.

The remaining pretest studies were designed to test the correctness of our intuitions concerning the relative codability of the characters' personalities in English and Chinese. In the first of these codability studies, 10 English monolinguals and 8 Chinese-English bilinguals were presented with the English and Chinese versions of the descriptions, respectively. They were then asked to label each character with the best overall personality trait or type term they could think of, and also to rate how “accurately and completely” their term captured the character's personality. They then gave their second choice of descriptive term for each character, also rating how well it described him.

For the second codability study, we first compiled a list of all English terms and a list of all Chinese terms (both first and second choices) used to describe any of the eight characters by subjects in the first study. To each of these lists we added the nearest equivalent, where one existed, of each term on the other list not already appearing in a fairly close translation, as well as a number of additional terms culled from subjects' responses in the attribute-listing study. This brought the total number of terms in each list to 103. A new group of English monolinguals (N = 20) and one of Chinese-English bilinguals (N = 19) were then presented with either the English or Chinese versions of the character descriptions and the trait list and were asked to (a) pick the one trait from the list that best described each character, (b) rate how accurately and completely that trait described him, and (c) select the second- and third-best trait descriptor for each character (ratings were not obtained for these traits).

Results

On the basis of the results of the two codability studies (to be described shortly), four of the eight characters—those developed to exemplify the two English and the two Chinese trait terms having no close equivalent in the other language—were eventually used in the main experiment to test the research hypotheses. The two personality types with economical labels in English but not in Chinese were the *artistic type* and the *liberal type*. The artistic character (David) was described with statements relating to the following categories of attributes (derived from the attribute-listing study described earlier): (a) artistic skills and interests, (b) the artist's cognitive style and temperament (intense, moody, imaginative, fantasy prone, etc.), and (c) the bohemian lifestyle (unconventional behavior, attitudes, possessions). The description of the liberal character (Richard) exemplified these attribute categories: (a) liberal (i.e., progressive, left-wing) social and political attitudes, (b) tolerance and open-mindedness, and (c) a humanitarian, people-oriented outlook. Although Chinese has separate terms for the three attribute categories comprising each of these two types, in neither case is there an economical expression that incorporates, in the way that the English terms *artistic type* and *liberal type* do, all three attribute categories (as well as others not directly represented in the character descriptions).

The two personality types with labels in Chinese but not in English were *shì gù* and *shēn cáng bú lòu*. The description of the shì gù character (Steven) exemplified these attribute categories: (a) greater than average experience of the world (well traveled, varied job experience, etc.), (b) a strong family orientation (in particular, devotion to the immediate family), and (c) well-developed social and interpersonal skills (especially the ability to smooth out difficult interpersonal situations). The shēn cáng bú lòu character (Kevin) was described with statements relating to the following attribute categories: (a) very knowledgeable and skilled in a wide variety of areas, both practical and intellectual, (b) reluctant to display this knowledge and skill unless it is absolutely necessary to do so, and (c) inconspicuous to the point of frequently being ignored or forgotten by others. Again, although English possesses labels for the individual attributes associated with these two personality types, it lacks economical terms capable of unifying these attributes into single, coherent concepts. The Chinese terms *shì gù* and *shēn cáng bú lòu,* on the other hand, imply exactly these two constellations of attributes (in addition to other attributes not explicitly represented in the character descriptions).

The data shown in Table 1 confirmed our intuitions regarding the differential codabilities of these four stimulus personalities in English versus Chinese. As can be seen, English-language subjects in both codability studies believed that their freely chosen descriptors for the two characters based on English-labeled schemas (artistic David and liberal Richard) were more accurate and complete than were their descriptors for the two characters based on Chinese-labeled schemas (shì gù Steven and shēn cáng bú lòu Kevin). Chinese-language subjects in both studies showed just the opposite pattern of ratings. Because the results of the two studies did not differ significantly (all F ratios for interaction effects involving study as a variable were less than 1), they were analyzed together to increase power (see the bottom panel of Table 1 for the combined means). This analysis revealed that the predicted Language of Processing × Language of Schema interaction was reliable, $F(1, 53) = 6.73$, $p < .02$, and that no other effects reached significance. In sum, then, these studies established that the four characters created to exemplify schemas having labels in one of the languages but not in the other are indeed more readily and accurately codable by speakers of that language.

Table 1
Subjects' Ratings of Their Freely Chosen Trait Descriptors in the Two Codability Studies

Language of processing	Language of schema: English[a]	Language of schema: Chinese[b]
Study 1		
English	6.95	6.12
Chinese	6.44	6.56
Study 2		
English	8.02	7.75
Chinese	7.79	8.47
Studies 1 and 2 combined		
English	7.49	6.94
Chinese	7.11	7.52

Note. In Study 1, ratings were made on scales of 1 to 9. In Study 2, ratings were made on scales of 1 to 11. The two studies were combined in an unweighted-means analysis of variance.
[a] The data in this column are averaged over the two characters based on English-labeled schemas.
[b] The data in this column are averaged over the two characters based on Chinese-labeled schemas.

The data of the second codability study were also examined for evidence of the predicted trait labelings, and the results generally confirmed that the characters were labeled in the intended manner. The label most frequently chosen for Steven by the Chinese-language subjects was, as predicted, *shì gù*, which was listed by the majority of subjects. The English-language subjects' most frequent choice of label for Steven was *considerate*, which, as the reader who will refer to the description of that character can verify, does not adequately capture the important elements of his personality. Also as predicted, the label most frequently chosen for Kevin by the Chinese-language subjects was *shēn cáng bú lòu.* listed by a large majority of subjects. The English-language subjects most often chose *reserved*, which omits the skill/knowledgeability component of this character's personality, among other things.

The label most frequently selected for David by the English-language subjects was, as predicted, *artistic*, listed by a clear majority of subjects. Chinese-language subjects most often chose *yì shù tīān fèn*, which means *artistically talented*, particularly in the visual or musical arts. This term, however, refers narrowly to artistic skill and carries no stereotypic connotations regarding the person's temperament or lifestyle, unlike the English term *artistic type*. In the case of Richard, the English-language subjects apparently regarded *open-minded* as a somewhat better descriptor than *liberal* (we had the latter label in mind when constructing this character), but *liberal* was the second most frequently listed term. Chinese-language subjects most often listed *si xiˇang xin cháo*, which roughly translates as *trendy* and is not a particularly apt description of Richard's personality.

In general, we found the trait-labeling data quite encouraging in view of the fact that subjects had 103 trait terms from which to select in describing the characters. After taking into account these and other data from the pretest studies, we reduced each character description from 18 to 15 statements in length (deleting what in our judgment was the weakest statement in each cluster). We also made a few final adjustments to some of the English and Chinese wordings, and the descriptions were considered ready for use in the main experiment.

Main Experiment

Method

Twelve English monolinguals and 24 Chinese-English bilinguals, all undergraduates at the University of Alberta, Canada, participated in the experiment. The following three groups comprised the between-subjects variable in this study: (a) English monolinguals performing the experiment in English (hereinafter referred to as the *E-E* group), (b) Chinese-English bilinguals performing the experiment in English (the *CE-E* group), and (c) Chinese-English bilinguals performing the experiment in Chinese (the *CE-C* group). Bilingual subjects had previously been asked to rate, on a scale of 1 to 9, their fluency in both English and Chinese. Only those giving themselves a rating of 5 or greater for both languages were eligible to participate. The bilinguals were randomly assigned to the CE-E and CE-C conditions. There were 7 women and 5 men in each of the three groups. Each subject participated in two sessions conducted 5 days apart. The same bilingual experimenter conducted both sessions for subjects in all three groups.

In the E-E and CE-E conditions, all instructions (written and oral), stimulus materials, and response forms were in English, and subjects in the CE-E condition were asked to make every effort to think (and respond) exclusively in English during the experiment. Chinese instructions, stimulus materials, and response forms were used in the CE-C condition, and subjects were asked to think and respond exclusively in Chinese.

In the first session, subjects were given a booklet containing the character descriptions, with instructions to read them as many times as necessary to form a distinct impression of each character's personality and to remember each character's name. Subjects were told that they would be required to answer questions about each character's personality when they returned 5 days later, but they were led to believe that they would not be expected to remember the descriptions in literal detail. The order of character descriptions was counterbalanced across subjects.

In the second session, subjects performed the following four tasks in the order listed:

Free impressions

The subjects were given a list of the characters' names and were asked simply to write down their impressions of each character's personality. They were allowed 3 min per character in which to complete this task.

The English- and Chinese-language free-impression data were each coded by two independent judges. One bilingual judge coded both the English- and Chinese-language responses. A second bilingual judge coded only the Chinese-language responses, and an English monolingual judge coded only the English-language responses. The first step in coding the data was to segment the responses into information units, the smallest units of independent, meaningful information. For example, a sentence that described a character as “shy but intelligent” would be segmented into two information units: shy and intelligent. Whenever there was a subsidiary phrase that was meaningless without the main phrase, the two were counted together as one unit. Segmentation agreement was 85% for the English-language data and 92% for the Chinese-language data. Disagreements were decided by another judge.

The segmented impression units were then classified into three categories: (a) description-based items—those referring to attributes implicit in the original character description; (b) schema-congruent items—those referring to attributes contained in the personality schema on which the character was based but not

directly represented in the character description (the attribute-listing data collected in the first pretest study served to define the contents of the personality schemas); for example, the attribute *unreliable* was not implicit in the description of David but is part of the *artistic type* schema and would therefore be scored as a schema-congruent item; and (c) miscellaneous items—this was a residual category for items referring to attributes not represented in the character description and not contained in the schema on which the character was based.

Classification agreement was 92% for the English-language data and 78% for the Chinese-language data. Disagreements were again decided by another judge. Preliminary inspection of these data indicated that, although our hypotheses had been strongly confirmed, the Chinese-language responses had apparently been scored according to a looser criterion than had the English-language responses. In particular, there was a higher overall proportion of schema-congruent classifications and a lower overall proportion of miscellaneous classifications in the Chinese-language data than in the English-language data. Therefore, the Chinese data were given to a new bilingual judge and the English data to a new monolingual judge. The Chinese-language judge was instructed to review the original classifications and to apply a stricter criterion for the schema-congruent category, that is, to reclassify borderline cases of schema-congruent items as miscellaneous items. The English-language judge was instructed to apply a looser criterion for the schema-congruent category, that is, to reclassify borderline cases of miscellaneous items as schema-congruent items. These revisions resulted in roughly comparable overall proportions of description-based, schema-congruent, and miscellaneous items in the Chinese- and English-language data.

Free recall

Subjects were told to write down as many statements from the original descriptions as they could remember. They were given 4 min per character in which to complete this task.

The English- and Chinese-language recall data were each rated by two independent judges. The judges first decided if a given item could be matched with one of the stimulus sentences; if so, it was given a score of either 1 (the judge was barely able to match the recalled item with the original item), 2 (a fairly good reproduction, but one containing significant omissions, additions, or distortions), or 3 (an excellent reproduction except for minor details such as names, locations, etc.). Interjudge agreement was 82% for the English-language responses and 90% for the Chinese-language responses. Disagreements were decided by another judge.

Recognition memory

Of these 12 statements, 6 were taken from the original description (2 from each of the three attribute clusters) and 6 were new statements. Three of the new statements (schema-congruent foils) depicted behaviors consistent with the personality type on which the character was based. The other 3 new statements (schema-irrelevant foils) depicted behaviors unrelated to either the original character description or to the schema.

Subjects were presented with 12 statements about each character and were asked to assign each statement a rating of either 1 (“I am certain this statement was not in the description”), 2 (“I think this statement was not in the description, but I'm not sure”), 3 (“I think this statement was in the description, but I'm not sure”), or 4 (“I am certain this statement was in the description”). Subjects had a total of 12 min in which to complete this task.

Inference

Subjects were shown six entirely new statements pertaining to each character and were asked to rate the likelihood that each statement would be true of the character. Ratings were made on a 13-point scale ranging from 1 (*very unlikely*) to 13 (*very likely*). Three of the statements (description-based items) were based directly on the original character description; that is, they described attributes implicit in the stimulus materials. An example of a description-based item pertaining to the artistic character is, "David is more easily moved by music, art, and literature than is the average person." The other three statements (schema-congruent items) described attributes that were not implicit in the character description but were congruent with the schema on which the character was based. An example is, "David drinks heavily at times and likes to try out hallucinogenic drugs." Subjects were allowed a total of 12 min in which to complete this task.

Hypotheses

We advanced the following hypotheses, all of which take the form of a Language of Processing × Language of Schema interaction:

1. Subjects who had processed the stimulus information in English (E-E and CE-E groups) would have greater difficulty distinguishing presented and nonpresented items on the recognition-memory test for the characters based on English-labeled schemas, whereas subjects who had processed the stimulus information in Chinese (CE-C group) would have greater difficulty for the characters based on Chinese-labeled schemas. This effect was predicted to be particularly evident in the recognition judgments made of the schema-congruent foils.

2. Subjects who had processed the stimulus information in English would recall more of the material presented about the characters based on English-labeled schemas, whereas subjects who had processed the stimulus information in Chinese would do so for the characters based on Chinese-labeled schemas.

3. Subjects who had processed the stimulus information in English would generate more schema-congruent items in the free-impression task, and assign higher likelihood ratings to the schema-congruent items on the inference test, for the characters based on English-labeled schemas, whereas subjects who had processed the stimulus information in Chinese would do so for the characters based on Chinese-labeled schemas.

Results

Hypothesis 1, which pertains to the recognition-memory test, was partially supported. Table 2 presents subjects' mean scores on the measure d', which is an index of overall sensitivity to old versus new items. There was, first of all, a nearly significant but theoretically uninteresting main effect of language of schema, $F(1, 33) = 2.96$, $p < .10$, indicating that subjects tended to be less accurate for the characters based on Chinese-labeled schemas (mean $d' = 2.20$) than for the characters based on English-labeled schemas (mean $d' = 2.49$). In addition to this main effect of character type, however, a contrast testing the predicted Language of Processing × Language of Schema interaction was also significant, $t(33) = 2.47$, $p < .01$. The CE-C group was much less accurate for the characters based on Chinese-labeled schemas, the CE-E group was only slightly less accurate for those characters, and the E-E group was slightly less accurate for the characters based on English-labeled schemas. A further contrast comparing the two English-language conditions did not approach significance ($t < 1$).

Table 2
Mean d' Scores From the Recognition-Memory Test

		Language of schema	
Language of processing	Group	English[a]	Chinese[b]
English	E–E	1.91	2.11
	CE–E	2.43	2.24
Chinese	CE–C	3.13	2.25

Note. E-E = English monolinguals using English. CE-E = Chinese–English bilinguals using English. CE–C = Chinese–English bilinguals using Chinese.
[a] The data in this column are averaged over the two characters based on English-labeled schemas.
[b] The data in this column are averaged over the two characters based on Chinese-labeled schemas.

Surprisingly, however, these differences in accuracy were due more to differences in the ability to recognize old items of information than to differences in the tendency to give false alarms to schema-congruent foils. Table 3 presents subjects' mean recognition ratings of previously presented items. The Language of Processing × Language of Schema interaction was significant, $t(3) = 2.37$, $p < .05$. As the table shows, both groups of subjects working in English less confidently recognized information previously presented about the characters based on English-labeled schemas, whereas subjects working in Chinese less confidently recognized information previously presented about the characters based on Chinese-labeled schemas.

Table 3
Mean Ratings of Previously Seen Recognition-Memory Items

		Language of schema	
Language of processing	Group	English[a]	Chinese[b]
English	E–E	3.25	3.42
	CE–E	3.32	3.52
Chinese	CE–C	3.54	3.46

Note. E–E = English monolinguals using English. CE–E = Chinese–English bilinguals using English. CE–C = Chinese–English bilinguals using Chinese.
[a] The data in this column are averaged over the two characters based on English-labeled schemas.
[b] The data in this column are averaged over the two characters based on Chinese-labeled schemas.

As it turned out, between-condition variability in the tendency to give false alarms to schema-congruent foils was very highly correlated with between-condition variability in the tendency to give false alarms to schema-irrelevant foils (the correlation between the mean recognition ratings given to the two types of foils across the six cells of the design was $r = .87$). Because of this, it would not be appropriate to test the second part of Hypothesis 1 (which pertains to schema-congruent false alarms) directly on the schema-congruent false-alarm scores per se. Therefore a derived measure was constructed, namely the subject's mean rating of the schema-congruent foils minus his or her mean rating of the schema-irrelevant foils. This measure provided an index of the subject's tendency to falsely recognize schema-congruent material

while simultaneously controlling for any tendency to give false alarms to all new items. The data for this relative false-alarm index are shown in Table 4. The hypothesized Language of Processing × Language of Schema interaction failed to reach significance, $t(33) = 1.43$, $.05 < p < .10$, although the pattern of means is more or less consistent with predictions (the E-E group showed a greater tendency to give false alarms for the characters based on English schemas, the CE-E group showed an equal tendency to give false alarms for both sets of characters, and the CE-C group showed a greater tendency to give false alarms for the characters based on Chinese-labeled schemas). The contrast comparing the two English-language conditions did not approach significance ($t < 1$).

Table 4
Mean Scores on the Recognition-Memory False-Alarm Index

		Language of schema	
Language of processing	Group	English[a]	Chinese[b]
English	E–E	.76	.62
	CE–E	.71	.71
Chinese	CE–C	.68	.93

Note. E–E = English monolinguals using English. CE–E = Chinese–English bilinguals using English. CE–C = Chinese–English bilinguals using Chinese. The index reported in this table is the subject's mean rating of the schema-congruent foils minus his or her mean rating of the schema-irrelevant foils.
[a] The data in this column are averaged over the two characters based on English-labeled schemas.
[b] The data in this column are averaged over the two characters based on Chinese-labeled schemas.

In retrospect, we believe that the weakness of the false-alarm results stemmed from the nature of the schema-congruent foils themselves. In general, these foils tended to describe behaviors that were roughly similar to behaviors appearing in the original character descriptions. This in turn may have acted to minimize any effect of schematic processing on subjects' tendency to give false alarms to these foils. Had we used schema-congruent foils that were more clearly distinct from the originally presented information, we suspect that the Language of Processing × Language of Schema interaction would have been stronger.

Hypothesis 2 (that there would be a Language of Processing × Language of Schema interaction on the amount of freely recalled information about the characters) was not at all supported. In fact, the free-recall data showed no significant effects whatsoever on the number of items recalled, the accuracy with which items were recalled, or any other measure that we derived from these data. On the basis of this result, along with the finding that subjects less confidently recognized old information pertaining to characters based on labeled schemas in their language of processing, we may definitely conclude that schematic processing had no facilitative effect on memory for previously presented information in this study.

Table 5 and Table 6 present the mean numbers of schema-congruent items in subjects' free impressions and mean ratings of the schema-congruent items on the inference test, respectively. Hypothesis 3 was strongly supported by both sets of results: Subjects' impressions contained more schema-congruent information, and subjects made stronger inferences regarding schema-congruent attributes, when the language in which they processed a stimulus character provided a label for the schema that the character exemplifies. The Language of Processing × Language of Schema interaction was significant both for the free-impression data, $t(33) = 4.09$, $p < .001$, and for the inference data, $t(33) = 2.72$, $p < .01$. The contrast

comparing the two English-language conditions did not reach significance in either analysis, $t = 1.38$ and $t < 1$, respectively. (The inference-test data also showed a main effect of language of schema in that subjects tended to make stronger inferences about the characters based on Chinese-labeled schemas than about the characters based on English-labeled schemas. Again, however, this is a theoretically insignificant result that merely shows that the schema-congruent items written for the Chinese-schema characters happened to be somewhat more strongly linked to their personalities than the items written for the English-schema characters were linked to theirs. The important finding is that the tendency to make stronger inferences about the Chinese-schema characters was much more pronounced in the case of Chinese-language subjects.)

Table 5
Mean Numbers of Schema-Congruent Items in Subjects' Free Impressions

		Language of schema	
Language of processing	Group	English[a]	Chinese[b]
English	E–E	1.50	0.38
	CE–E	0.92	0.38
Chinese	CE–C	0.67	1.33

Note. E–E = English monolinguals using English. CE–E = Chinese–English bilinguals using English. CE–C = Chinese–English bilinguals using Chinese.
[a] The data in this column are averaged over the two characters based on English-labeled schemas.
[b] The data in this column are averaged over the two characters based on Chinese-labeled schemas.

Table 6
Mean Ratings of Schema-Congruent Attributes on the Inference Test

		Language of schema	
Language of processing	Group	English[a]	Chinese[b]
English	E–E	8.33	9.01
	CE–E	8.15	9.06
Chinese	CE–C	7.53	10.01

Note. E–E = English monolinguals using English. CE–E = Chinese–English bilinguals using English. CE–C = Chinese–English bilinguals using Chinese.
[a] The data in this column are averaged over the two characters based on English-labeled schemas.
[b] The data in this column are averaged over the two characters based on Chinese-labeled schemas.

In contrast to the results just described, there was no Language of Processing × Language of Schema interaction in the case of description-based attributes, either in the free-impression data ($t < 1$) or in the inference data ($t < 1$).

Discussion

This study tested whether languages with differing repertories of labeled personality-type schemas are capable of exerting language-specific effects on their speakers' memory for and impressions of other people. It succeeded in demonstrating that both impressions and memory are affected when the target's personality and behavior conform to a labeled schema in the perceiver's language. These results should provide a broader, cross-linguistic perspective on current schema research in addition to extending the growing body of work on the language-cognition relation into the domain of social cognition (cf. Hoffman, Mischel, & Baer, 1984).

An interesting pattern of results emerged from the free-impression and inference tasks. There were significant Language of Processing × Language of Schema interactions on the number and strength of schema-congruent inferences but not of description-based inferences. Thus, ready access to an appropriate labeled schema facilitated "going beyond the information given" but did not act to strengthen impressions or predictions concerning attributes implicit in the targets' actual behavior.

The recognition-memory data provide further evidence of linguistic effects on thinking and also suggest a possible explanation for the pattern of results obtained with the free-impression and inference tasks. To restate the primary findings, subjects whose language of processing labels the schema on which a given character was based were less confident in recognizing previously presented items of information about the character and also tended (although nonsignificantly) to have greater difficulty in rejecting nonpresented but schema-congruent items of information. These findings suggest that the availability of a labeled schema may have led subjects to focus relatively less on the details of a target's behavior and instead rely more on their schematic knowledge of his personality type. Such a strategy has the advantage of greatly reducing the cognitive effort required of the perceiver, but also has the obvious disadvantage of causing confusion in memory between presented and nonpresented information.

On the other hand, subjects without the benefit of a labeled schema apparently attended more closely to the target's behavior, as they exhibited superior recognition memory for previously presented information. This analysis can also explain the lack of a Language of Processing × Language of Schema interaction on description-based responses in the free-impression and inference tasks. Inferences about stimulus-based attributes could potentially have been reached in two ways—by consulting the appropriate schema or by consulting one's memory for the stimulus information—and the two srategies apparently yielded comparable results. In contrast, it is only by way of a schema that one can transcend the direct implications of the presented information, which accounts for the significant Language of Processing × Language of Schema interaction on schema-congruent responses in these tasks.

It is of some interest to compare our findings with the results of the research on language and color memory, which represents the most extensive body of work to date on the language-and-thought question. Except for the substitution of personality types for colors, the logic of our study was very similar to that of the color-memory experiments—particularly the study by Stefflre et al. (1966), in which an actual cross-linguistic comparison was made. In both cases, the research strategy involved looking for possible effects of the linguistic codability of stimuli on memory for the stimuli (and, in our study, on inferences about the stimuli as well). The color-memory research, however, has consistently shown that the availability of an appropriate verbal description results in more accurate recognition memory for a given color, whereas our study found that ready access to an appropriate labeled schema resulted in less accurate recognition memory for a target person's behavior.

These divergent findings may have stemmed in part from differences in the recognition-memory tasks used. In our study, subjects were required to recognize facts about the stimulus person, whereas in the color-memory research the task was simply to identify the stimulus in an array. Had our study also required subjects to identify the target person in a group of similar individuals, it is possible that a facilitative effect of schema availability would have been found. Intuitively, however, this seems unlikely. We tend to think that the differences in results stem from differences in the nature of the stimuli themselves. Many of the important attributes of social stimuli—personalities in particular—are hidden from the perceiver's view, and consequently there is a need, or at least the opportunity, to rely heavily on the inference process in fleshing out one's representation of the stimulus. This, of course, can lead to confusion in memory between observed and inferred attributes—all the more so to the extent that one can apply a schematic label to the stimulus. In the case of colors, there is obviously little if any need or even possibility for the perceiver to infer additional attributes of the stimulus, if indeed it makes sense to speak of colors as having attributes at all. Consequently, if one's descriptive label for a color is a reasonably good one in the first place, there is no reason to expect it to bias one's later memory for the color.

In many respects, color stimuli may be the exception rather than the rule. Certainly it seems to be true that most other nonsocial stimuli, including the natural objects studied by cognitive psychologists, more closely resemble most social stimuli than they do colors. There is, however, at least one fairly obvious difference in the effects of schematic processing on the representations that people form of social and nonsocial stimuli. This difference has to do not so much with the extent of people's inferences about the two types of stimuli but more with the accuracy of those inferences. In most cases where a natural object can be labeled with a relatively brief verbal expression at all, the label constitutes a highly useful description of the object, in the sense of being highly predictive of the object's attributes. Inferences about the important features of objects categorized as chairs, trees, or automobiles are very likely to be correct. In contrast, inferences about the attributes of persons categorized as artistic types or extraverts are nearly as likely to be wrong as to be right, if we accept the current wisdom regarding the generality and consistency of personality (or rather the lack thereof). People could pay a considerable price in accuracy by attempting to apply the same type of categorical thinking that serves them so well in the object domain to the world of social experience.

We believe that our study provides somewhat more conclusive evidence for an effect of language on thought than have most previous studies using a cross-linguistic design. We say this because the design of this study allowed us to evaluate the role of nonlinguistic cultural differences between the groups as possible causes of the obtained effects, which is not the case with studies that simply contrast monolingual speakers of different languages or, more generally, with studies that limit their comparisons to two groups of subjects with differing cultural characteristics. Because the E-E and CE-E groups for the most part responded similarly but differed from the CE-C group, we have clear evidence for an effect of language per se. It is true, however, that although in no case did the difference between the E-E and CE-E groups reach significance, on some measures the CE-E group did tend to fall between the E-E and CE-C groups. This suggests that cultural differences also played a role in our results, albeit a smaller one than did the language-of-processing variable.

On the other hand, there are at least two respects in which this study furnishes only limited evidence in support of the linguistic relativity hypothesis, beyond the obvious fact that the results are consistent only with the weak Whorfian view (i.e., that a given language makes certain ways of thinking either easier or more difficult) and not with the strong view (i.e., that a given language makes certain ways of thinking either obligatory or impossible). In the first place, we are unsure if the effects of language demonstrated here generally characterize most realworld cases of person cognition. Subjects in our study received relatively impoverished input about a relatively large number of stimulus persons and had to retain this information over a fairly long period of time. It is precisely under these difficult memory conditions that

we would expect the effects of verbally labeled schemas to be most pronounced. When people are able to focus their attention in a more leisurely manner on the behavior of individual persons, their impressions may not be structured to the same degree, if at all, in terms of labeled schemas such as *artistic type*. In the second place, this study has demonstrated only that a language's repertory of labeled categories (its lexicon) affects the categorizing behavior of its speakers. Far more controversial, and also far more interesting, are the possible effects of a language's grammar on the thought patterns of its speakers. Whorf himself believed that a language's grammar embodies the linguistic community's world view. With few exceptions, however (e.g., Bloom, 1981; Carroll & Casagrande, 1958), the effects of grammar on thought and behavior have gone unstudied, probably because it is so difficult to operationalize and measure the sorts of effects hypothesized by Whorf. The possibility that these fascinating ideas will someday yield to empirical scrutiny nonetheless remains an exciting one.

Footnotes

[1] The two characters exemplifying trait terms with essentially identical meanings in the two languages were omitted from the main study to reduce subjects' memory load and because their inclusion was not essential to evaluate the hypotheses under study. The two characters exemplifying trait terms with a missing attribute in one language did not seem to be perceived in the intended manner by subjects in the pretest studies. Although modified versions of these two characters were included in the main study for exploratory purposes, they yielded no interpretable results and will therefore not be discussed further in this article.

[2] Chinese terms are given here in pīn yīn romanization kindly provided for us by J. S. Lin of the Department of East Asian Languages and Literatures, University of Alberta. Chinese-language materials in the actual studies were, of course, written in Chinese ideographs.

[3] The test for the Language of Processing × Language of Schema interaction, in this and all other analyses, was as follows. Because language of schema is a within-subjects variable, a difference score was first calculated for each subject by subtracting the mean score for the two characters based on Chinese-labeled schemas from the mean score for the two characters based on English-labeled schemas. The hypothesized interaction could then be tested by applying the following contrast to the difference scores just described: 1 (E-E), 1 (CE-E), −2 (CE-C). Because a directional pattern of means was predicted, one-tailed tests of significance were used.

[4] The E-E and CE-E groups were compared by applying the following contrast to the difference scores described in Footnote 3: 1 (E-E), −1 (CE-E). Because no differences between these two groups were anticipated, two-tailed tests of significance were used.

[5] Of relevance here is the distinction that Lingle, Altom, and Medin (1984) drew between *membership attributes* and *inference attributes*. Membership attributes are those that are (or can be) used to assign instances to categories in the first place, whereas inference attributes are those are simply implied by the fact of category membership. According to Lingle et al., membership attributes and inference attributes tend to be the same for natural object categories but not for social categories, particularly person categories. If this is true, it provides an obvious explanation for the difference in inferential accuracy that, as we have argued, is probably associated with the two types of categories.

References

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.

Bloom, A. H. (1981). *The linguistic shaping of thought: A study in the impact of language on thinking in China and the West*. Hillsdale, NJ: Erlbaum.

Brislin, R. W. (1970). Back-translation for cross-cultural research.. *Journal of Cross-Cultural Psychology*, *1*, 185–216.

Brown, R., & Lenneberg, E. (1954). A study in language and cognition.. *Journal of Abnormal and Social Psychology*, *49*, 454–462.

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.

Carmichael, L., Hogan, H. P., & Walter, A. A. (1932). An experimental study of the effect of language on the reproduction of visually perceived form.. *Journal of Experimental Pschology*, *15*, 73–86.

Carroll, J. B., & Casagrande, J. B. (1958). The function of language classifications in behavior. In E. E.Maccoby, T. M.Newcomb, & E. L.Hartley (Eds.), *Readings in social psychology* (3rd ed., (pp. 18–31). New York: Holt, Rinehart & Winston.

Clark, H. H. (1969). Linguistic processes in deductive reasoning.. *Psychological Review*, *76*, 387–404.

Fiske, S. T., & Taylor, S. E. (1984). *Social cognition*. Reading, MA: Addison-Wesley.

Forgas, J. P. (1983). The effects of prototypicality and cultural salience on perceptions of people.. *Journal of Research in Personality*, *17*, 153–173.

Greenfield, P. M., Reich, L. C., & Olver, R. R. (1966). On culture and equivalence: II. In J. S.Bruner, R. R.Olver, & P. M.Greenfield (Eds.), *Studies in cognitive growth* (pp. 270–318). New York: Wiley.

Higgins, E. T., & Chaires, W. M. (1980). Accessibility of interrelational constructs: Implications for stimulus encoding and creativity.. *Journal of Experimental Social Psychology*, *16*, 348–361.

Hoffman, C., Mischel, W., & Baer, J. (1984). Language and person cognition: Effects of communicative set on trait attribution.. *Journal of Personality and Social Psychology*, *46*, 1029–1043.

Lantz, D., & Stefflre, V. (1964). Language and cognition revisited.. *Journal of Abnormal and Social Psychology*, *69*, 472–481.

Lingle, J. H., Altom, M. W., & Medin, D. L. (1984). Of cabbages and kings: Assessing the extendibility of natural object concept models to social things. In R. S.Wyer, Jr., & T. K.Srull (Eds.), *Handbook of social cognition* (*Vol. 1*, pp. 71–117). Hillsdale, NJ: Erlbaum.

Lucy, J. A., & Shweder, R. A. (1979). Whorf and his critics: Linguistic and nonlinguistic influences on color memory.. *American Anthropologist, 81*, 581–615.

Rosch, E. (1974). Linguistic relativity. In A.Silverstein (Ed.), *Human communication: Theoretical explorations* (pp. 95–121). New York: Halstead Press.

Stefflre, V., Castillo Vales, V., & Morley, L. (1966). Language and cognition in Yucatan: A cross-cultural replication.. *Journal of Personality and Social Psychology, 4*, 112–115.

Whorf, B. L. (1956). *Language, thought, and reality*. New York: Wiley.