WHAT IS THE SAPIR-WHORF HYPOTHESIS?

By

Paul Kay
University of California, Berkeley

and

Willett Kempton
Michigan State University

Cognitive Science Program

Institute of Cognitive Studies

University of California at Berkeley

April 1983

Acknowledgement: The authors have received help on this material from a number of colleagues too great to list. We would like to single out Brent Berlin and Eleanor Rosch for special thanks. This research was supported by grant BNS 78-15900 from the National Science Foundation. The assistance of Don Burgess with the Tarahumara field trials is gratefully acknowledged.

The production of the Berkeley Cognitive Science Report Series is supported by a grant in cognitive science to the University of California at Berkeley from the Alfred P. Sloan Foundation.

Historical Background

The doctrine of radical linguistic relativity is to be understood historically as a reaction to the denigrating attitude toward unwritten languages that was fostered by the evolutionary view prevalent in anthropology in the nineteenth century. Subsequently, the research of Boas and his students showed these languages to be as systematic and as logically rich as any European language, and it was perhaps inevitable that the latter finding should spawn a doctrine on non-European languages and cultures antithetical to the evolutionary view. If this doctrine of radical relativity has led to certain excesses of its own--in which the valid insistence on the integrity of each linguistic system has led to an underestimation of their common structural features--we should not forget that it nonetheless supplied a needed corrective to the ethnocentric evolutionism it replaced. Indeed, outside of certain rarified academic milieux, the early relativists' battle for a rational and unprejudiced view of our nonliterate contemporaries is not yet won.

Sapir was Boas' student and Whorf Sapir's. The following two passages are among the most frequently cited from Sapir and Whorf, respectively, and the reader's indulgence is begged that we print them yet again. In the first, Sapir expresses, in terms of no less lucid for being poetic, the basic empirical finding of the Boasians on the formal completeness and intellectual adequacy of unwritten languages.

> Both simple and complex types of language of an indefinite number of varieties may by found spoken at any desired level of cultural advance. When it comes to linguistic form, Plato walks with the Macedonian swineherd, Confucius with the headhunting savage of Assam (Sapir 1921:219).

In the second passage, Whorf takes the further step, foreshadowed in other writings of Sapir's (*e.g.*, 'The "real world" is to a large extent unconsciously built up on the language habits of the group' 1951 [1929]:160), of claiming that an intellectual system embodied in each language shapes the thought of its speakers in a quite general way.

> 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 have to be organized in our minds. This means, largely, by the linguistic system in our minds (Whorf 1956 [1940]:212f).

During the past forty years, there have been many attempts to recast the fundamental insights of Sapir and Whorf, originally expressed in a number of evocative and sometimes metaphorical passages similar to those just cited, in terms of sufficiently prosaic that the doctrine may be subjected to empirical test. We will not attempt to review that literature but rather endorse Roger Brown's conviction that Eric Lenneberg in 1953 really said all that was necessary (Brown 1976:128). In Brown's summary, 'Whorf appeared to put forward two hypotheses:

(1) Structural differences between language systems will, in general, be paralleled by nonlinguistic cognitive differences, of an unspecified sort, in the native speakers of the two languages.

(2) The structure of anyone's native language strongly influences or fully determines the world-view he will acquire as he learns the language' (Brown 1976:128).

It seems that Whorf's interpreters have imbued his doctrine with an additional tacit premise. If the differences in world-view mentioned in (2) and the non-linguistic cognitive differences mentioned in (1) are to be interesting, they must be sizeable. Miniscule differences are dull. But if the differences in the dependent variable (non-linguistic cognition) are big, then probably the differences in the independent variable (language) are also big. (This last is not a logical consequence, but appears to be a piece of tacit, plausible reasoning that has in fact occurred historically.) Hence (1) and (2) appear to have induced the tacit postulation of (3) on the part of Whorf's followers.

(3) The semantic systems of different languages vary without constraint.

Traditions of Empirical Research

Empirical research inspired by the Sapir-Whorf hypotheses may be divided into two, largely independent traditions, one devoted to the evaluation of (1) and the other to the evaluation of (3). Until a technique is developed for assessing the world-view of a people independently of the language they speak, no direct test of (2) is possible. It seems probably that in attempting to find empirical support for (3), anthropologists have sought to provide indirect evidence for (2).

The bulk of the research in both traditions has concerned the domain of color. Empirical work on color devoted to evaluation of the Sapir-Whorf hypothesis by anthropologists and linguists belongs mainly to the tradition concerned with (3). Studies in this tradition have sought to describe and compare the semantics of the color lexicons of different languages. Some of the better known, early studies are those of Gleason (1951), Ray (1952), Conklin (1955), Lenneberg and Roberts (1956), Bohannon (1963), and Berlin and Kay (1969). (From the early seventies on, there have been a large number of studies in this tradition.) The studies prior to 1969 tended to support (3); those since 1969 have tended to discredit (3). Kay and McDaniel (1978) summarize the findings of the later studies regarding constraints operating in color classification across languages and attempt to explain these constraints as a consequence of the neurophysiology of human color vision. In this paper we accept the main lines of Kay and McDaniel's conclusions and so take (3) to be discredited. There appear to be strong constraints on possible interlinguistic variation in the encoding of color.

The experiments reported in this paper belong to the tradition of research, primarily conducted by psychologists, concerned with evaluating (1). We should note, however, before closing this section, that since empirical work on the Sapir-Whorf hypothesis has been restricted essentially to the domain of color, the above conclusions are correspondingly restricted. There are other areas of human thought and belief--religion is an obvious example--in which constraints like those imposed by peripheral neural mechanisms on possible color classifications seem *a priori* unlikely to operate. Such domains therefore offer greater potential scope for application of notions like (2) and (3).

The basic research paradigm in the evaluation of hypothesis (1) has involved establishing a correlation between a linguistic and a non-linguistic, cognitive variable within a single language. Among the defining studies in this tradition of research are those of Brown and Lenneberg (1954), Burnham and Clark (1955), Lenneberg (1961), Lantz and Stefflre (1964), and Stefflre, Castillo and Morely (1966). The linguistic variable has come in two forms, 'codability' and 'communication accuracy.' The non-linguistic variable has been 'memorability.' Support for the Sapir-Whorf hypothesis was claimed when a correlation was found between the memorability of a color and its value on one of the linguistic variables. Rosch (Heider 1972; Heider and Olivier 1972) has shown, however, that focality or perceptual salience universally determines both memorability and codability/communication accuracy. Prior to Rosch's work, codability and communication accuracy were *assumed* to differ across languages and once this is shown to be false the correlation obtaining between memorability and a linguistic variable no longer gives support to the Whorfian thesis. (See Brown 1976 for a lucid exposition of this and related points, especially pp. 113f, 149.)

Recently, Lucy and Shweder (1979) have performed experiments that they interpret as reinstating the Whorfian claims made by the early studies. Lucy and Shweder succeed in developing a stimulus array that exhibits a communication-accuracy/memorability correlation independent of focality. However, their way of doing this is to begin with a standard stimulus array and selectively substitute chips into this array until the result has the desired property. Thus, Lucy and Shweder claim to have shown that a stimulus array can be constructed in which there is a communication-accuracy/memorability correlation independent of focality.

What such a correlation may mean is, however, unclear. ‘Communication accuracy’ is established with two groups of subjects. Members of the first group are shown an array of colors and instructed to name each color so that another person could pick the color out. The second group is given the array of colors and the set of names produced by the first group and asked to pick out the colors intended by the various names. Colors that are more often picked out correctly are defined as having higher communication accuracy. Communication accuracy is therefore a function not only of the code embodied in the language used but also of the particular array of contrasts comprised by the stimulus set. For example, a focal green stimulus will be highly codable as *green* in any array. However, its communication accuracy score will be high in an array lacking other very green chips but low in an array that contains other very green chips. Since the Lucy-Shweder stimulus array is purposely constructed by successive approximations until it yields a (focality-independent) communication-accuracy/memorability correlation, it is unclear what meaning the Lucy-Shweder correlation may have outside of the particular array of colors that was deliberately constructed to produce it.

However one judges the claims and counterclaims in this research tradition, it seems clear that any within-language correlational study, matching a linguistic variable purporting to measure the degree to which the language ‘favors’ a color to a nonlinguistic cognitive variable like memorability, is at best an indirect test of (1). Any such study depends on quite tenuous inference to connect the experiment with the theory it is supposed to evaluate. We are now in a position to perform more direct tests of (1).

Design of Experiment I

A direct test of the Sapir-Whorf hypothesis (1) will involve the following: for the linguistic variable, we select subjects from two languages that differ in color terminology. For example, English makes a basic lexical distinction (Berlin and Kay 1969:5ff) between the color categories ‘green’ and ‘blue’ while Tarahumara, a Uto-Aztecan language of Northern Mexico, lacks this basic lexical distinction, having instead the basic term *siyóname*, which means ‘green or blue’ (Burgess, Kempton and MacLaury 1983). The non-linguistic, cognitive variable will be subjective distance between colors, that is, subjective judgments of how similar or different colors are. The question is whether the linguistic difference--*green/blue* vs. *siyóname*--will produce a difference in subjective distance between colors. The Sapir-Whorf hypothesis, as expressed in (1), predicts that colors near the *green-blue* boundary will be subjectively pushed apart by English speakers precisely because English has the words *green* and *blue*, while Tarahumara speakers. lacking this lexical distinction, will show no comparable distortion.

Before describing the experiment, two explanatory preliminaries are in order. The first concerns the ‘real’ psychological distances between colors, and the second concerns the ‘boundary’ between color categories. To say that English speakers may ‘distort’ the distances between colors close to the *green-blue* boundary is to presuppose some ‘real’ scale of psychological distance for color (which is then subject to the distortion). The scale of psychological distance between colors we take as the ‘real’ scale for present purposes is called discrimination distance. The unit of this scale is the just noticeable difference (jnd), that is, the smallest physical difference in wavelength that can be detected by the human eye. The jnd is a variable (and nonmonotonic) function of wavelength because the eye is sensitive to finer differences in wavelength near the center of the visible range than at the edges. Whereas wavelength measures the physical basis of the psychophysical color response, the jnd scale is the direct measure of that response.

When we speak of the ‘boundary’ between lexical categories, this boundary is defined, not in absolute terms, but as the point of exact balance of two overlapping, gradient categories. In particular, the *blue-green* lexical category boundary is defined as that wavelength at which an equal mixture of green and blue is perceived. The *blue/green* boundary was established empirically as follows. Four American English subjects were exposed to a standard series of forty hues at two levels of brightness and asked for each color to rate the degree of blueness, greenness, redness and yellowness on a scale of 0 to 9. For each chip and each color word the ratings were averaged. The results yielded smooth curves with the topological features of those given in Kay and McDaniel (l978:624ff) and in Figure 1. The *blue/green* lexical category boundary was defined as the point of intersection of the *blue* and *green* curves, as indicated in Figure 1.

Figure 1

Figure 2

Stimuli

The experimental stimuli consisted of eight color chips of varying shades of greenish blue and blueish green and representing two levels of (medium) brightness (Munsell levels 5 and 6). Figure 2 shows the relation of these stimuli to each other in terms both of the discrimination distances among them and their location relative to the English *blue-green* lexical category boundary. The vertical dotted line in Figure 2 depicts the lexical category boundary; the fact that this line passes through chip G indicates that chip G represents an equal mixture of green and blue (i.e., G is a perfect aqua or turquoise).

The numeral appearing between each pair of letters denotes the discrimination distance between the chips indicated by the letters. For example, pairs (B,C) and (C,D) are separated by equal discrimination distances, while the discrimination distance between chips A and B is 1.27 times as great as the (B,C) distance. All discrimination distances are normalized to (i.e., expressed as proportions of) the smallest (B,C) discrimination distance. The color chips are from the Munsell collection and the discrimination distances are computed from the tables of Wyszecki and Stiles (1967:450-500), which is the computational method recommended by the CIE Committee on Colorimetry.

To forestall a plausible but spurious objection that might otherwise distract the reader, the discrimination distance norms were established on English speaking subjects, to be sure, but since the only subjects found in the experiment to deviate from these norms are also the English speakers, the absence of discrimination data from Tarahumara speakers is irrelevant. The Munsell notations for the stimuli shown in Figure 1 are as follows: A: 7.5G 6/10 B: 2.5BG 6/8 C: 7.5BG 6/8 D: 2.5B 6/8 E: 10.0G 5/10 F: 5.0BG 5/8 G: 10.0BG 5/8 H: 5.0B 5/8.

Measuring the Dependent Variable: Judged Distance

Probably the simplest way to elicit subjective distance among *n* stimuli (here *n* = 8) is the triad technique. In this technique the experimenter presents three items (a 'triad') from the set of *n* stimuli, and asks the subject which of the three is most different from the other two. This is repeated for each possible triad. In the present case there are eight color stimuli, A,B,...,H, and therefore 56 possible triads: (A,B,C), (A,B,D),...,(F,G,H). Presentation order of triads and of stimuli within each triad were randomized for each interview.

Every pair of stimuli, e.g. (A,B), will occur in *n-2* triads (here 6). The psychological distance between A and B relative to other stimulus pairs in the set is given by the proportion of times A and B are split by the subject's selection of one of them as the most different item in the triad. For example, if in every triad in which A and B occur the third color (C , D, E, F, G or H) is always chosen as most different, then A and B are minimally distant; conversely, if either A or B is always chosen, then A and B are maximally distant. In general, the proportion of times two stimuli are split up in selecting the odd member of the triad indicates their distance relative to other pairs of stimuli. We are only interested here in ordinal comparisons of distances, so relative distances suffice.

It is well known that triads distances for color stimuli reproduce discrimination distances faithfully, so long as the stimuli selected are chosen from a single lexical category--e.g., a single hue of red varied by saturation and brightness (Torgerson 1958; Messick 1954). The classic Torgerson result has been widely taken to establish that triads distances represent a general measure of physiologically possible discrimination distances, regardless of lexical categories. We will see, however, that this is not the case.

Experiment I

Experiment I consisted in the administration of the full set of 56 triads to each of five color-normal English and four color-normal Tarahumara speakers and the calculation of the mean inter-chip relative distances for each pair of chips adjacent on the hue scale in each group separately. The Tarahumara language was used for interviews of the Tarahumara speakers, which took place in their native village in Chihuahua, Mexico.

The results are summarized in Table 1, where the triads distances obtained from the English and Tarahumara speaking subjects are compared with each other and with the nature of the stimuli as regards both discrimination distances and lexical boundary location. For each pair of triads distances compared, the numbers shown are normalized on the smaller difference of that pair. Hence in each pair of Tarahumara-English triads distances, the smaller is always given as '1.00.'

Table 1

In the upper left cell of Table 1 we find the comparison of the (A,B) and (B,C) distances. It may be noted that chips B and C are closer to each other in discrimination distance than are chips A and B, while the *blue-green* lexical category boundary passes between chips B and C. We see that, while the Tarahumara put a bigger triads distance between A and B than between B and C--which agrees with the discrimination distances--the English speakers exaggerate the (B,C) subjective (triad) distance, making it bigger than the (A,B) distance. This is the distortion predicted by the Sapir-Whorf hypothesis (1): exaggeration of subjective differences that cross a lexical category boundary.

The second (upper right) block of Table 1, comprising the (E,F) and (F,G) triads distances, has the same logic as the block just considered, and shows the same result. The (F,G) pair is adjacent to the lexical category boundary, but has the smaller discrimination distance than the (E,F) pair. The Whorfian prediction is that the Tarahumara speakers will judge E more distant from F than F is from G, following the discrimination distances, while English speakers will make the opposite choice, biased by the lexical category boundary to ignore the greater discrimination distance between E and F. Once again the numbers bear out the Whorfian prediction.

In block three (lower left), comprising stimuli B, C, and D, the logic of the test is slightly more complicated. Here there is no difference in discrimination distance between the two pairs of chips. The Whorfian prediction is that the English speakers should, with no difference in discrimination distance to overcome, place a massively greater triads distance between chips B and C than between C and D, reflecting the fact that the lexical category boundary passes between B and C. The data confirm this prediction. But we note that the Tarahumara also judge the colors separated by the *English* lexical boundary (B,C) as more distant than the other pair (C,D). On reflection, this latter result, though not deducible from Whorfian principles, is not surprising. As Kay and McDaniel (1978) point out, the categories ‘blue’ and ‘green’ (and also ‘red,’ ‘yellow,’ ‘black,’ and ‘white’) are covertly present in those languages which do not accord them a basic color term, as these categories correspond to the six categories of fundamental neural response for colors identified by De Valois and his associates (e.g., De Valois, *et al.* 1968). Most languages that have a word ‘grue’ and no separate words for ‘blue’ and ‘green’ nevertheless contain fixed or semi-fixed expressions such as 'grue like the sky'/'grue like the leaves' which demonstrate that the concepts 'blue' and 'green' are present even if not named at the level of basic color terms. In the Kay and McDaniel formulation, in fact, composite categories (such as 'grue'), in languages that contain them, are defined as the union of more basic fundamental neural response categories ('green' and 'blue'). Thus the Kay-McDaniel formulation claims that these categories are present cognitively for speakers of all languages, whether or not the language accords them a separate word, and so accounts for the effect under discussion. This conclusion receives empirical support, specifically for Tarahumara green and blue, in Burgess, Kempton and MacLaury (1983). We note that this lexically covert distinction affects the Tarahumara triads distance only when the discrimination distances are precisely equal.

Finally, in the fourth block of Table 1, since the central stimulus G is directly on the lexical category boundary and the discrimination distances are virtually identical, no test of the Whorfian effect is possible. Tarahumara and English speakers each give approximately equal triads distances.

A test of the statistical significance of the overall Whorfian effect was made. Each subject's triads distance for each pair of hue-adjacent chips can be scored as greater or less than what would be predicted by the discrimination distance between this pair of chips. (The predicted triads distance for a pair is the product of the mean triads distance for all pairs and the ratio of the discrimination distance on this pair to the mean discrimination distance for all pairs.) Thus each of the six hue-adjacent triads distances for each subject provides a separate test of the hypothesis. The five English subjects are predicted by the Sapir-Whorf hypothesis to 'stretch' distances (B,C), (F,G), and (G,H) and to 'shrink' distances (A,B), (C,D), and (E,F). Of the thirty subject by chip-pair predictions of stretch or shrink, twenty-nine were vindicated in the data. Taking the thirty separate predictions as independent binomial trials (which is technically not quite legitimate), if the true probability of success on a single trial were 1/2, a result this favorable--29 or more wins in 30 trials--would occur by chance less than once in thirty million repetitions.

By contrast, the Tarahumara speakers, for whom random behavior with respect to stretching and shrinking is predicted, vindicated the hypothesis of Whorfian effect 13 times in 24 trials, nearly a perfect 50-50 split.

Discussion of Experiment I

Experiment I shows a Whorfian effect. The presence of the *blue-green* lexical category boundary appears to cause speakers of English to exaggerate the subjective distances of colors close to this boundary. Tarahumara, which does not lexicalize the blue-green contrast, does not show this distorting effect.

What psychological mechanism might explain this result? We note first that it cannot be the case that the vision of English speakers is distorted in some way by the language they speak, because the discrimination distances that the Tarahumara faithfully reproduce on the subjective triads task were established on speakers of English. What we have to explain is why English speakers' subjective judgments of distance between colors near a lexical category boundary departs from the visual (discrimination) distances.

What then accounts for the subjective judgments of the English speakers? We first construct a frankly speculative kind of 'Just-So Story.' Although this story is not initially supported by independent evidence, it generates a second experiment whose results are consistent with it.

The stimuli employed in Experiment I were similar enough to each other that intuitively a triad like (A,B,C) presented a difficult judgment to make. When one is shown the triad (A,B,C), it is obvious only that B is not the most different. We propose that faced with this situation the English speaking subject reasons unconsciously as follows: 'It's hard to decide here which looks the most different from B, A or C. Are there any other kinds of clues I might use? Aha! A and B are both CALLED *green* while C is CALLED *blue*. That solves my problem; I'll pick C as most different.' Of course this cognitive strategy, which we will call the 'name strategy,' is not available to the Tarahumara speaker precisely because he doesn't have ready lexical labels for the concepts green or blue.

Although we shall pursue the argument for the name strategy, we have nothing specific to say about the cognitive level at which it may operate. When we present the crucial triples of colors to sophisticated English speaking subjects, describe the name strategy to them, and ask them to suppress any tendency to use this strategy, they make the same subjective judgments as naive subjects; further, they report that the chip selected just 'looks' the most different, regardless of the name. Thus if the name strategy operates for English speakers, it operates at a level that is not only out of consciousness but which cannot easily be brought under conscious control.[1]

To test whether unconscious employment of the name strategy is in fact the psychological mechanism underlying the Whorfian effect discovered in Experiment I, we want to design an experiment in which the name strategy cannot be used and see whether the Whorfian effect disappears. In such an experiment, English speaking subjects' judgments of subjective distance should, like the Tarahumaras', be predictable from discrimination distance and show no sensitivity to lexical category boundaries.

Experiment II

The second experiment uses the same stimuli as the first. Here also stimuli are presented three at a time; the subject is, moreover, asked for a judgment logically equivalent to that asked in the first experment--pick an 'odd man out' of the three chips. However, the method of presentation and instructions insure that the task is not psychologically equivalent.

In Experiment II each of the four triads of chips representing two hue-adjacent pairs--(A,B,C), (E,F,G), (B,C,D), and (F,G, H)--were presented to each of twenty-one color-normal English speaking subjects in the following manner. The three chips were arranged in a container with a sliding top that permitted the subject to see alternately either of two pairs of the three chips, but never all three at once. For example, in triad (A,B,C) the pairs alternately made visible were (A,B) and (B,C). The method of presentation and instructions were as follows:

> Experimenter exposes pair (A,B). "You can see that this chip (points to A) is greener than this chip (points to B)." (All subjects readily agreed.) Experimenter slides cover so that A is covered and C exposed along with B; i.e., the pair (B,C) is now exposed, "You can see that this chip (points to C) is bluer than this chip (points to B)." (Again all subjects agreed without problems.) "Now," experimenter hands stimuli to subject, "you may slide the cover back and forth as often as you like. I'd like you to tell me which is bigger: the difference in greenness

[1] Those readers who have Munsell chips available and whose dominant language distinguishes green and blue can try this experiment for themselves. We recommend triples (A,B,C) and (E,F,G); see the text for Munsell specifications of the chips. Tarahumara speakers systematically pick chips A and E as most different. English speakers, even those warned to try to suppress the name strategy, pick chips C and G as most different.

> between the two chips on the left or the difference in blueness between the two chips on the right."

In this display, the chip placed centrally in the box, and which is also intermediate in hue between the other two (e.g., chip B), is in view of the subject at all times. It is obvious that there are just three chips involved in the presentation. Subjects do not experience the presentation or instructions as tricky or hard to understand.

Although the judgment asked of the subject in Experiment II is formally the same as that asked in Experiment I--i.e., to pick out the most different chip from the three presented (e.g., A,B,C)--Experiment II is designed to block use of the name strategy. The subject cannot reasonably ask himself whether chip B is called *green* or *blue* because he has already in effect both called it *green* and called it *blue* in agreeing to compare B in *greeness* to A and in *blueness* to C. It is thus irrelevant to this task whether chip B would be called *green* or *blue* in another, ‘neutral’ context.

Results of Experiment II

In Experiment II, the Whorfian effect shown by English speaking subjects in Experiment I disappears. Subjective similarity judgments follow discrimination distance and reflect no influence from lexical category boundaries. These results are summarized in Table 2. Numbers tabulated are simply the number out of 21 subjects who chose the indicated pairwise subjective distance as larger. Thus in the first block, 17 out of the 21 subjects picked the (A,B) greenness difference as larger than the (B,C) blueness difference--in effect saying that A is the most different chip in the triad (A,B,C). This agrees with the discrimination distances and with the Tarahumara judgments in Experiment I and disagrees with the Whorfian prediction and with the English judgments in Experiment I.

Table 2

In block 2 of Table 2 (chips E,F,G) this result is repeated exactly. Again English speakers reverse their performance of Experiment I, locating the bigger subjective difference--by a vote of 17 to 4--at the larger discrimination distance rather than at the lexical category boundary.

In the third block of Table 2 (chips B,C,D) there is no difference in discrimination distance between the two pairs, although the lexical category boundary splits chips B and C. If the task is, as we intended, highly sensitive to discrimination distance and insensitive to lexical category boundaries, the prediction is for no difference in this block. Although there is a slight difference, it is not significant at the .05 level.[2]

In the final block of Table 2 (chips F,G,H), where there is a small discrimination distance difference and no category boundary bias (because the boundary goes through the central chip G), we find the expected discrimination distance effect.

To sum up the results of Experiment II, when we alter the task from that of Experiment I so that the same stimuli are being judged with respect to the subjective distances among them but the name strategy is blocked, the lexical category boundary (Whorfian) effect is no longer able to override discrimination distance.

Discussion

In the introductory sections of this paper we cited R. Brown's summary of E. Lenneberg's exposition of the Sapir-Whorf hypothesis (1-2), which we repeat for the reader's convenience.

(1) Structural differences between language systems will, in general, be paralleled by nonlinguistic cognitive differences, of an unspecified sort, in the native speakers of the two languages.

(2) The structure of anyone's native language strongly influences or fully determines the world view he will acquire as he learns the language.

We suggested that a third hypothesis has been tacitly assumed, because otherwise the claim that (1) makes is undramatic.

---

[2] If in a larger sample the 2/1 ratio were maintained in favor of placing the larger subjective distance where the lexical category boundary falls, then we would be able to speak of a mild category boundary effect on this task in the absence of discrimination distance conflict.

(3) The semantic systems of different languages vary without constraint.

Hypotheses (3) and (1) have each spawned a tradition of empirical research on color. With respect to the anthropological tradition, we believe that recent work in the color domain has shown (3) to be wrong. (See Kay and McDaniel 1978.) The present study has therefore been devoted to the evaluation of a more limited kind of Whorfianism--a less dramatic linguistic determinism in which linguistic differences may be capable of inducing cognitive differences, but the nature and detailed effects of such differences are taken as open empirical questions.

Our first experiment demonstrated a Whorfian effect regarding subjective judgments of similarity among colors. It was found that in direct subjective judgment of distance (the triads task) English speakers' judgments are distorted in the region of the *blue-green* lexical category boundary, while the subjective judgments of speakers of Tarahumara--which lacks the *blue-green* lexical distinction--show no such distortion. We noted that this effect on English speakers is obviously not caused by an aberration of vision, since the discrimination norms were themselves established on English speakers.

We were led therefore to seek a more cognitive explanation for the effect and proposed the name strategy. According to the name strategy hypothesis, the speaker who is confronted with a difficult task of classificatory judgment may use the *lexical* classification of the judged objects as if it were correlated with the required dimension of judgment even when it is not, so long as the structure of the task does not block this possibility. Thus, when we are asked to pick out the most different-looking of three chips in the blue-green region, if it is not obvious which of the extreme chips is farthest from the middle chip in appearance, we may substitute the nonequivalent but more easily solved problem of asking ourselves which of the extreme stimuli has a different *name* from the middle stimulus. (We noted, however, that the name strategy seems to operate at an utterly nonconscious level, since sophisticated subjects to whom the name strategy is described nevertheless report that the color in question 'looks different'.)

The second experiment confirmed the name strategy interpretation. The experimental setup precluded the subject's use of the name strategy, and the Whorfian effect found in Experiment I disappeared; English speakers' subjective similarity judgments revert to conformity with the underlying discrimination distances.

We have considered some evidence supporting the 'name strategy' hypothesis as the explanation of the effect that appears in Experiment I and disappears in Experiment II. Let us assume for the moment that further research establishes the name hypothesis as empirically correct and ask how it is related to the Sapir-Whorf hypothesis. If the name hypothesis is right, are Sapir and Whorf supported? The answer would seem to depend a lot on how one interprets Sapir and Whorf, and their writings are notoriously subject to multiple interpretations. Certainly, if we take the radical linguistic determinist side of Whorf, illustrated in the 'kaleidoscopic flux of impressions' passage cited in the introduction to this paper, the name strategy does not support this kind of Whorfianism. The name strategy does not agree with *radical* linguistic determinism, in which the structure of the language imposes its categories as the only categories in which we can experience the world, or in which to use Sapir's words, 'human beings...are very much at the mercy of the particular language' (1959 [1929]:160). This is so precisely because the distortion of English judgments can be made to disappear, proving that we are not hopelessly at the mercy of our language. On the other hand, the name strategy does seem to support the kind of modest Whorfianism expressed in (1): between English and Tarahumara speakers we have a difference in nonlinguistic cognition (subjective judgments of similarity) correlated to a difference in linguistic structure. One could quibble about whether it is accurate to call use of the name strategy 'nonlinguistic' cognition, but then one could quibble on the other side over whether the word 'nonlinguistic' in (1) is really essential to the Sapir-Whorf position.

Both Sapir and Whorf were wont to suggest in certain sweeping and dramatic passages a kind of absolute linguistic determinism and linguistic relativity that elsewhere each qualified considerably. There are qualifications and hedges in the writing of both that suggest neither believed in the total linguistic relativism and determinism that seems sometimes to speak from their pages and which has been greatly emphasized by their more ardent, if not their most careful, devotees. For example, Sapir could have made an absolute claim about the tyranny of language, rather than a hedged claim, by omitting the words ‘very much’ in the passage cited just above.

In a similar vein, Whorf wrote extensively of *habitual* thought and behavior and its relation to language. What exactly Whorf meant by the word ‘habitual’ is hard to decide from its various uses in his work, especially in his famous paper where the word figures in the title (1969 [1941]). It is possible to give Whorf‘s use of the word ‘habitual’ a construal that fits quite well with our name strategy hypothesis, including especially the fact that the name strategy can be suppressed.

In developing his notion of *habitual* thought and behavior and how it differs between English and Hopi speakers, Whorf first considers lexical and grammatical facts about the treatment of time in English. He says, ‘In our language...plurality and cardinal numbers are applied in two ways: to real plurals and imaginary plurals. Or more exactly if less tersely: perceptible spatial aggregates and metaphorical aggregates: We say "ten men" and also "ten days"’ (1956 [1941]:139). The claim here is that some things really are plural (or really are experienced directly as plural) while other things have the conceptual structure of plurality imposed on them by a metaphor that in another language could be and often is avoided. Everybody, Whorf seems to be saying, has to experience ten men as an aggregate, but we English speakers extend this aggregate schema to days, while the Hopi do not. A few lines later Whorf again suggests that he conceives experience as having two tiers; one, a kind of rock bottom, inescapable seeing-things-as-they-are (or at least as human beings cannot help but see them) and a second, in which the metaphors implicit in the grammatical and lexical structures of language cause us to classify things in ways that could be otherwise (and are otherwise for speakers of different languages). Thus Whorf continues, ‘Our *awareness* of time and cyclicity does contain something immediate and subjective--the basic sense of "becoming later and later." But in the habitual thought of us SAE [e.g., English speaking] people, this is covered under something quite different...’ (1956 [1941]:139). Again, Whorf seems to be saying that we share with the Hopi an *awareness* [his emphasis] of the getting-later-and-later aspect of our concept of time, apparently because this is just how it is, or at least how humans necessarily experience it. But part of our concept of time is not shared with the Hopi--its quantification into periods (hence its pluralizability). Whorf is of course emphasizing the nonnecessary, metaphorical part of our experience, but he is equally presupposing the necessary, awareness part, and apparently presupposing further that the awareness part is universal, either prelinguistic or panlinguistic or both. As some careful commentators on Sapir and Whorf have pointed out, they couldn't really have believed the absolute linguistic relativity they sometimes appeared to profess and at the same time believed they could explain in English what, for example, the Hopi view of time is. This comment is particularly relevant to the passage of Whorf's under discussion. He seems to say that we English speakers 'habitually' think of time as a segregable and aggregable quantity (or perhaps tend to think of it this way), but this habit can be broken. That is, our habitual thought *tends* to be fitted to a particular linguistic mold; in fact, Whorf often implies that we not only can but should break this mold. This is neither absolute linguistic relativism nor absolute linguistic determinism.

A more cautious Whorfianism seems to be supported by the results reported here and by other contemporary research on color. In this view we acknowledge that there are constraints on semantic differences between languages, so we accept not an absolute linguistic relativity but a modest version. Statement (3) is rejected and we recognize that (2), having always suffered from unclarity, is now reduced in its consequences as well. Regarding (1) we note that there do appear to be incursions of linguistic categorization into apparently nonlinguistic processes of thinking, even incursions that result in judgments which differ from those made on a purely perceptual basis. Thus, employing the name strategy the English speaker judges chip B to be more similar to A than to C because the *blue-green* boundary passes between B and C, even though B is perceptually closer to C than to A. The name strategy seems to demand two facilitating conditions: (1) it must not be blocked by the context as in Experiment II; (2) the original judgment must be in some sense hard to make.[3]

It is possible to give Sapir and Whorf readings that accord with this empirically motivated view of limited linguistic relativity and determinism. Such a reading is not the one usually given

[3] In order to detect a Whorfian effect using triad tests, it was necessary to choose stimulus pairs with small differences in discrimination distance; *e.g.*, if chip A had been perceptually far from chip B, the large difference in discrimination distance would surely have overcome the lexical category boundary effect, and the English speaking subjects would have picked A as the most distant of the three chips. We were not, therefore, able to test the possibility that a Whorfian effect might be found even if the triads judgments were easy to make.

and is certainly not what most anthropology students are taught as 'The Sapir-Whorf Hypothesis.' What either Sapir or Whorf actually believed on this topic is of course impossible to know, especially as the writings of both men are open to such varied interpretations. The question of what these two scholars thought, although interesting, is after all less important than the issue of what is the case. The case seems to be first, that languages differ semantically but not without constraint and second, that linguistic differences may induce non-linguistic cognitive differences but not so absolutely that universal cognitive processes cannot be recovered under appropriate contextual conditions.

References

Berlin, Brent and Paul Kay.

1969 *Basic Color Terms*. Berkeley: California.

Bohannon, Paul.

1963 *Social Anthropology*. New York: Holt, Rinehard & Winston.

Brown, Roger and E. H. Lenneberg.

1954 A study in language and cognition. *Journal of Abnormal and Social Psychology* 49:454-62.

Brown Roger.

1976 Reference: In memorial tribute to Eric Lenneberg. *Cognition* 4:125-53.

Burgess, Don, Willett Kempton and Robert E. MacLaury.

1983 Tarahumara Color Modifiers: Category Structure Presaging Evolutionary Change. *American Ethnologist*. In press.

Burnham, R. W. and U. R. Clark.

1955 A test of hue memory. *Journal of Applied Psychology* 39:164-72.

Conklin, H. A.

1955 Hanunóo color categories. *Southwestern Journal of Anthropology* 11:339-44.

De Valois, R. L., I. Abramov, and G. H. Jacobs.

1966 Analysis of response patterns of LGN cells. *Journal of the Optical Society of America* 56:966-77.

Gleason, H. A.

1955 *Introduction to Descriptive Linguistics*. New York: Henry Holt.

Heider, E. R.

1972 Universals in color naming and memory. *Journal of Experimental Psychology* 93:10-20.

Heider, E. R. and C. C. Olivier.

1972 The structure of the color space in naming and memory for two languages. *Cognitive Psychology* 3:337-54.

Kay, Paul and Chad K. McDaniel.

1978 The linguistic significance of the meaning of basic color term. *Language* 54:610-46.

Lantz, D. and V. Stefflre.

1964 Language and cognition revisited. *Journal of Abnormal and Social Psychology* 69:472-81.

Lenneberg, E. H. and John R. Roberts.

1956 The language of experience; A study in methodology. *International Journal of American Linguistics* (Memoir No. 13).

Lenneberg, E. H.

1961 Color naming, color recognition, color discrimination: A reappraisal. *Perceptual and Motor Skills* 12:375-82.

Lucy, John and Richard Shweder.

1979 Whorf and his critics; Linguistic and nonlinguistic influences on color memory. *American Anthropologist* 81:581-615.

Messick, S. J.

1954 *The perception of attitude relationships: A multidimensional approach to the structuring of social attitudes*. Ph.D. Thesis. Princeton.

Ray, Verne F.

1952 Techniques and problems in the study of human color perception. *Southwestern Journal of Anthropology* 8:251-59.

Sapir, Edward.

1921 *Language*. New York: Harcourt, Brace & World.

1951 [1929] The status of linguistics as a science. *Selected Writings*. David Mandelbaum, ed. Berkeley: California, [*Language* 5:207-14].

Stefflre, V., V. Castillo and L. Moreley.

1966 Language and cognition in Yucatan; A cross-cultural replication. *Journal of Personality and Social Psychology* 4:112-115.

Torgerson, Warren S.

1958 *Theory and Methods of Scaling*. New York: Wiley.

Whorf, Benjamin Lee.

1956 [1940] Science and Linguistics. *Language, Thought and Reality*. John B. Carroll, ed. M.I.T.: Cambridge [*Technology Review* 42:229-31, 247-48].

Whorf, Benjamin Lee.

1956 [1941] The relation of habitual thought and behavior to language. *Language, Thought and Reality*. John B. Carroll, ed. M.I.T.: Cambridge. [*Language, Culture and Personality, Essays in Memory of Edward Sapir*. Leslie Spier, ed. Menasha, Wisc.: Sapir Memorial Publications Fund].

Wyszecki, Gunter and W. S. Stiles.

1967 *Color Science, Concepts and Methods, Quantitative Data and Formulas*. New York: Wiley.