Research Article

A SEMANTIC SPACE OF COLOR NAMES

Ch.A. Izmailov and E.N. Sokolov
Moscow State University

Abstract—*Three Russian subjects learned arbitrary pairings between 20 colors and 20 three-letter artificial color names. After different amounts of this training, the subjects rated the difference between the colors associated with every pair of artificial names when these names were presented without the colors. Multidimensional scaling of the ratings after a small amount of training revealed a grouping of the words into four semantic clusters corresponding to the following groups of related colors: the violets, the blues, the greens, and the yellows-through-reds. After more extensive training, multidimensional scaling yielded the full color circle of hues. Further analysis of the data indicated that a spherical model previously proposed by the authors for sensory color space has advantages, also, for the semantic color space obtained when only the names of colors are presented. The results are interpreted in terms of a two-stage process of neuronal analysis of visual inputs in which the activity of four color-opponent channels is followed by differential activation of cells tuned to specific colors.*

The construction of a semantic space of color names attracts scientists' attention as much as the construction of a subjective space of color stimuli. The results of applying various methods, such as the semantic differential, factor analysis, and multidimensional scaling, can be summed up as follows:

1. The scaling of basic color names, mostly those relating to the basic colors of the spectrum (e.g., *yellow, green,* and *red*), yields a two-dimensional space of color names similar to Newton's color circle (Fillenbaum & Rapoport, 1971; Shmelev, 1983).
2. The inclusion of color names that refer to material characteristics of colored objects (e.g., *lemon, emerald, marsh-green, khaki, coffee,* and *gold*) increases the dimensionality and disorganizes the circular structure of the basic color space (Artemieva, 1968; Sokolov & Vartanov, 1987).

The main reason for the discrepancy between the results of scaling these two types of color names—the basic and the material names—lies hypothetically in the difference of their semantic content and structure. The semantics of a basic color name is determined by the corresponding sensory experience offered by an individual's visual system. The semantics of a material color name includes, besides visual experience, some other types of cognitive experience, especially that of speech (Artemieva, 1968; Miller, 1971).

An experimental way of testing this hypothesis might be to create a set of artificial color names whose semantics is determined beforehand, and then compare the semantic space of these artificial color names with that of natural ones. This article presents the results of a series of experiments aimed at building such a semantic space of artificial color names.

METHOD

Apparatus and Stimuli

The stimuli were generated by two Elektronika-C420 color TV sets controlled by CM-1403 computer. Colors were displayed on one of the screens, words on the other. The colors were distributed over the spectrum between violet and red, and also included white and a number of purple mixtures of short and long wavelengths. The subjects' reactions were put into the computer through an 11-key miniterminal.

Subjects

The experiments were performed with three subjects, aged 22 to 25, with normal color vision; two subjects served in the main experiment and one only in an auxiliary experiment.

Procedure

The experimental procedure consisted of three stages. First, we picked out 20 three-letter words that have no color meaning for Russian-speaking subjects. The initial spatial structure of the chosen words was tested by metric multidimensional scaling (Izmailov, 1980; Torgerson, 1958). Our aim was to obtain the initial background structure of the meaningless words on the list. Second, each subject was then taught, by simple associative training, pairings between the words and 20 colors (Artemieva, 1968; Miller, 1969, 1971). Third, the same method of multidimensional scaling was used to construct a semantic space for the artificial color names. In this way, we obtained information on the changes in the initial semantic structure brought about by color experience.

Phase 1

In Phase 1, pairs of the artificial words from the main list were successively presented on the TV screen. For each pair, the subject estimated the difference in color between the words' meanings on a scale from 0 (complete identity) to 9 (maximum difference). Each word in a pair was shown for 0.8 s, with an

---

Address correspondence to Ch.A. Izmailov, Moscow State University, 18/5 Department of Psychology, 103009, Moscow, Prospekt Marxa, Russia.

interval between words of 0.4 s. Two seconds after the first pair, a second pair was presented, then a third, and so on, until every possible two-word combination of the 20 words on the list had appeared 10 times. The order of presentation was random.

For each subject, we obtained a triangular matrix of 20 (20 − 1)/2 elements, each element being the arithmetic mean of 10 estimates. Every matrix was processed by metric multidimensional scaling. We calculated the eigenvalues of the coordinates in a 20-dimensional euclidean space, and then the coordinates of the 20 points representing the 20 words. The axes were numbered in order of their eigenvalues. Subsequently, we calculated the linear correlation coefficients between the initial difference estimates and the interpoint distances in the 20 dimensions of subjective space. The results for the first 6 dimensions are summarized in Table 1.

Phase 2

The main purpose of this phase was to train the subjects to name each of the 20 colors by a single one of the 20 artificial words on the list. The colors were mixtures of three basic TV colors with the following dominant wavelengths: red 620 nm, green 535 nm, and blue 485 nm. The colors were mixed in pairs in such a way that the stimuli included all the colors of the spectrum plus purples. In addition to chromatic colors, white was obtained by mixing all three TV colors. All colors were equated in brightness to the white by the method of heterochromatic photometry at the level of 15 cd/m.

One of the 20 stimuli would appear on the screen of one of the TV displays. The corresponding word would appear on the screen of the other TV display with a slight delay. The color–word pairs were presented five times each in random order. Immediately after this training, the subjects were shown only the colors and asked to name each by its corresponding artificial word. If all the names were given correctly, the training was terminated. Otherwise, the training continued until the subject learned to give the assigned names to all the colors. Only then did we pass on to the third experimental phase—the building up of a semantic space of artificial color names.

Table 1. *Characteristic roots and coefficients of correlation obtained in the multidimensional scaling analysis of the matrices for two subjects in Phase 1, before training*

Dimension of space	Characteristic root		Coefficient of correlation	
	L.S.	L.Sh.	L.S.	L.Sh.
1	7667	13330	.52	.52
2	4118	9504	.64	.66
3	3195	8103	.66	.77
4	2783	5824	.70	.81
5	2610	4515	.72	.87
6	2449	3715	.77	.89

Note. Results for the first six dimensions only are presented.

Phase 3

The procedure used in the third phase was identical to that in the first phase. We again obtained a matrix of difference estimates for every pair of words and analyzed it by metric multidimensional scaling. The matrix of semantic differences is given in Table 2, and the resulting eigenvalues and correlation coefficients in Table 3.

RESULTS AND DISCUSSION[1]

The semantic space of artificial words before training appears random in its projection on the first two principal axes. Moreover, the decrease of the eigenvalues of the axes is relatively uniform, and so is the increase of the correlation coefficient with number of dimensions. This pattern indicates that the selected artificial words were initially devoid of color semantics.

The results of the multidimensional scaling of semantic differences after training offer an entirely different picture. There is a noticeable shift in the eigenvalues between the first two axes and the rest. This shift implies that the first two axes determine the location of the points in space, while the other dimensions are unimportant. The projections of the 20 points representing artificial color names onto the X_1X_2 plane of the euclidean space are given in Figure 1. The points are distributed around four loci at some distance from each other. Each locus comprises points representing the names of colors from one of the four different parts of the visual spectrum.

For example, in Figure 1a, the names of colors from the indigo-blue part of the spectrum (Points 7–9) are grouped at the bottom right, the names of green colors (Points 11–13) are on the bottom left, the names of yellow and orange colors (Points 14–19) are at the top left, and the names of violet colors (Points 2–5) are on the top right.

The points inside the loci are not ordered. The subject confused the names of neighboring colors, but not the names of colors from different loci. The fact that the artificial color names fall into four distinct classes based on wavelength permits us to relate the four natural color names—*red, blue, green,* and *yellow*—to Hering's four basic colors. We can see that the semantics of artificial color names is determined mainly by color stimulation. This conclusion holds for both subjects (although the results for each have some individual features).

The obtained configuration of the points differs considerably from the color circle. The question arose whether what we obtained is the final or an intermediate structure of the artificial color names. Therefore, we repeated the Phase 2 training a week later with the same subjects. This time the subjects were much quicker to learn all the words. Then, once again, we built a semantic space of color names for each subject (Table 4).

It is clear from the data that additional training did not change the dimensionality of the semantic space of artificial color names. It only increased the metric precision of the two-

---

[1] Tables of the following data may be obtained from the authors: matrix of semantic differences between stimulus words before learning associations with colors and matrix of semantic differences between stimulus words after a second session of training.

Table 2. *Semantic differences between artificial words obtained for two subjects after Phase 2 training*

Word	No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
ФИР	1		66	62	54	66	66	82	86	80	80	86	86	88	74	52	34	50	10	10	84
БУМ	2	34		24	16	18	16	76	78	74	82	82	86	82	84	68	68	68	68	46	84
ВАП	3	30	10		10	14	16	68	76	76	84	80	84	80	82	72	76	68	68	60	84
ГАЧ	4	48	56	48		12	14	76	74	70	76	78	82	84	78	66	76	70	64	70	62
ДАХ	5	58	46	54	10		10	76	74	70	76	80	80	82	80	68	74	72	60	64	72
ЖАД	6	78	58	64	34	36		74	72	68	68	82	82	80	80	80	80	74	64	74	76
КИБ	7	90	64	60	30	30	22		14	12	40	72	78	80	82	80	80	80	82	80	58
ЛУС	8	78	66	64	32	44	20	14		10	24	70	80	76	80	76	80	84	80	82	28
ДЭК	9	90	66	58	30	34	24	10	14		30	74	82	76	78	80	80	74	78	80	42
ХАЦ	10	74	68	62	36	42	18	28	28	24		72	76	68	80	78	82	84	82	80	44
ЖОК	11	78	60	74	68	70	70	72	72	74	34		28	12	76	78	80	80	78	82	76
ЗАН	12	90	78	84	80	80	86	88	82	86	74	28		10	82	82	78	82	84	78	80
МЕК	13	86	70	80	74	78	82	84	80	88	78	23	10		76	82	80	80	80	80	78
НЮЖ	14	58	56	70	62	66	70	76	70	72	44	34	68	70		28	26	46	62	68	76
ПЕВ	15	46	54	50	58	60	66	68	72	72	62	52	72	70	26		24	36	40	66	78
РУН	16	26	32	34	62	60	78	76	76	80	74	66	80	74	48	46		32	36	56	82
САВ	17	34	40	38	54	54	68	60	64	62	52	62	74	74	30	20	28		34	24	80
ТОЛ	18	10	32	36	62	58	76	82	76	62	72	74	78	82	54	38	30	22		10	84
БИФ	19	10	30	34	60	64	78	86	78	84	74	72	82	82	58	48	48	34	18		82
ДИЛ	20	60	54	54	56	54	44	50	48	48	28	34	56	58	30	42	48	46	54	60	

Note. Data were averaged across 10 presentations of each word pair. The upper right-hand triangle presents the data for subject L.S., and the lower left-hand triangle presents the data for subject L.Sh.

dimensional space of artificial color names. The difference between the eigenvalues of the first two axes and of all the other axes increased, also. The difference between the correlation coefficients for a two-dimensional space and those for spaces with more than two dimensions decreased. This result corroborates the claim that the space of artificial color names has only two dimensions.

The configuration of points representing the artificial color names in a two-dimensional euclidean space is shown in Figure 2a (for subject L.S.) and 2b (for subject L.Sh.). Comparison with Figure 1 gives an idea of the specific changes in the semantic space caused by additional training. The points previously inside each of the four groups of color names are now situated on the circumference of the color circle, in accord with the hues of the corresponding colors. This pattern is especially clear in the data in Figure 2a.

Table 3. *Characteristic roots and coefficients of correlation obtained in the multidimensional scaling analysis of the matrices in Table 2*

Dimension of space	Characteristic root		Coefficient of correlation	
	L.S.	L.Sh.	L.S.	L.Sh.
1	16389	15236	.66	.66
2	12496	10802	.86	.94
3	9328	4186	.94	.97
4	4277	1987	.96	.96
5	2420	1668	.98	.96
6	1653	1559	.98	.97

Note. Results for the first six dimensions only are presented.

These results suggest that subjects first learn to group new color words into general categories based, perhaps, on opponent-color channels, and then learn a refined organization of the words in agreement with the spectral arrangement of the color circle. Because the formation of the basic color categories is thus determined by the activity of the visual system, the semantic space of basic color names such as *blue*, *green*, *yellow*, and *white* corresponds closely to the sensory color space.

This conclusion can be interpreted more accurately if we analyze the results in terms of the spherical model of color vision (Izmailov, 1980, 1982; Izmailov & Sokolov, 1991; Sokolov & Vartanov, 1987). The reason is that the spherical model imposes more limitations on the structure of the sensory color space than traditional euclidean models of color vision.

The Spherical Model of Color Concept Space

The main idea underlying the spherical model is that in the visual system, light is analyzed by means of four neuronal channels: two chromatic channels (*red-green* and *blue-yellow*) and two achromatic channels (*bright* and *dark*). The channels' outputs become the inputs of the color-detector cells, each tuned selectively to a certain color determined by a specific combination of the coefficients of synaptic transmission. Although each color detector has its own combination of synaptic coefficients, the sum of the squares of the coefficients is constant across detectors. This corresponds mathematically to the equation of a spherical surface in four-dimensional euclidean space. A set of points on the surface of such a sphere represents the set

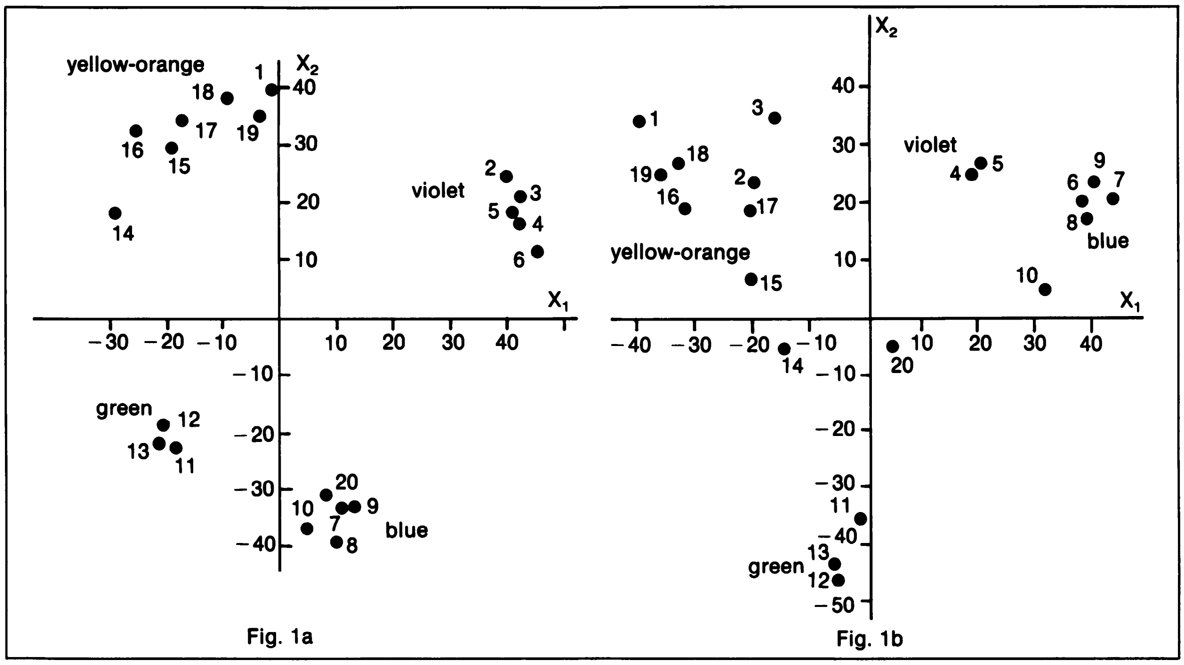

Fig. 1a

Fig. 1b

Fig. 1. Semantic space of artificial color names after one training session. Points corresponding to color categories form four loci in approximate accordance with Hering's opponent colors. (a) Subject L.S. (b) Subject L.Sh.

of colors discriminated by the visual system. Each color point is fixed by setting a single combination of four coordinates, although the same combination can correspond to lights with different spectral distributions (Izmailov, 1980, 1982; Izmailov, Sokolov, & Chernorizov, 1989).

The two chromatic channels analyze the spectral distribution of light that is perceived as hue. The achromatic channels analyze light intensity, perceived as color brightness. Because the chromatic and achromatic channels are related via the spherical law, the combination of their activities yields another color parameter—saturation, which compensates to a certain extent for the loss of physical information in visual processing. Adding saturation defines third polar coordinates of the four-dimensional color space.

Table 4. *Characteristic roots and coefficients of correlation obtained in the multidimensional scaling analysis of the matrices for two subjects after a second training session*

Dimension of space	Characteristic root		Coefficient of correlation	
	L.S.	L.Sh.	L.S.	L.Sh.
1	15766	15010	.72	.74
2	12715	13094	.89	.94
3	8360	3800	.95	.96
4	5086	2230	.97	.97
5	1693	1484	.98	.96
6	1354	1248	.98	.97

Note. Results for the first six dimensions only are presented.

The three characteristics—hue, brightness, and saturation—are represented in the spherical model as three spherical coordinates. The model gives a new description, on the one hand, to the interconnection of the neurophysiological channels, which are represented by the four Cartesian coordinates of a color point on the sphere, and, on the other hand, to the sensory characteristics of light, represented by the three spherical coordinates of the same point.

For equibright lights from different parts of the spectrum, as used in our experiments, the subset of colors falls on a spherical surface in a three-dimensional euclidean space. If the structure of artificial color names reflects the activity of the sensory system, then the configuration of artificial color names obtained by multidimensional scaling should best be represented on the surface of a sphere in a three-dimensional space, also, and not on a plane, as in Figures 1 and 2. If so, Figures 1 and 2 give only the projections of the points onto the X_1X_2 plane. A third dimension would be necessary to represent the configuration fully.

By testing the sphericity of the configurations, we can answer the following question: On which level of color information processing in the sensory system does the final shaping of basic color categories take place? For every multidimensional scaling in a three-dimensional euclidean space, we found the center of the sphere—the theoretical point as nearly as possible equidistant from all the points. Since the data contain error, the spherical layer has a certain *thickness,* defined as the standard deviation of the radii. On each iteration, a program calculates the radii of a sphere with a given center, the mean of the radii, and the standard deviation. The program then adjusts the position of the center until the standard deviation is minimum. For a three-dimensional space, the solution is considered acceptable if the thickness of the sphere containing the experimental

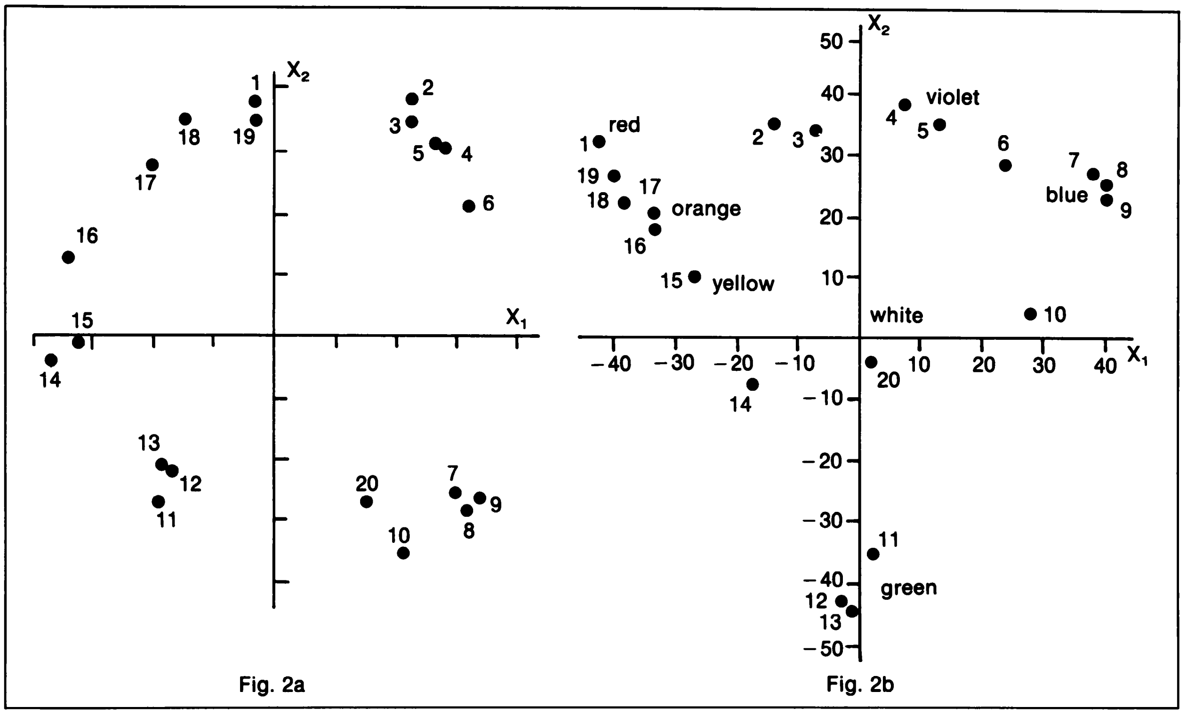

Fig. 2a

Fig. 2b

Fig. 2. Semantic space of artificial color names after a second session of training that fixed the results of the first run. The configuration resembles Newton's color circle, which is typical for colors in sensory space. (a) Subject L.S. (b) Subject L.Sh.

points does not exceed 10 to 12% of the mean radius (Izmailov, 1980; Izmailov et al., 1989).

The results of testing all the experimental configurations for sphericity are given in Table 5. The thicknesses of the layers containing the experimental points before training are 27% and 21%, respectively, for subjects L.S. and L.Sh. There is no significant sphericity. This is additional evidence of the absence of initial structure in the artificial words. Training brings about a decrease in the variation of the radii, reflecting the shaping of a spherical structure.

Additional Data and Conclusions

Color category formation in the course of training divides into two stages. The first stage results in the separation of artificial words into four classes perhaps corresponding to Hering's opponent colors. In the second stage, colors are differentiated in hue according to Newton's color circle. Additional evidence for this conclusion comes from the data of subject V.L., obtained after more extensive (five runs) learning of color names. The resulting multidimensional scaling solution for artificial color names, presented in Figure 3, has the same circular order of points as in Newton's circle, and the sphericity of this space is most like that found for sensory color space (Table 5; Izmailov, 1980; Izmailov et al., 1989; Sokolov & Vartanov, 1987).

An analysis of sphericity also supports the idea that there are two stages in the learning process. Color-opponent axes are formed at the first stage of training. The spherical structure of the semantic space of colors appears later, with repetition and fixing of word–color associations.

The activity of the sensory system can affect the brain areas

Table 5. *Indices of sphericity of three-dimensional semantic spaces in terms of radii of points representing artificial color names*

	Before learning		After 1 run		After 2 runs		After 5 runs
Index	L.S.	L.Sh.	L.S.	L.Sh.	L.S.	L.Sh.	V.L.
M	29.6	38.6	44.9	43.4	43.2	44.8	37.2
SD	8.0	8.1	2.7	7.0	2.3	6.1	4.2
CV	27.0	21.0	6.0	16.0	6.5	13.5	11.0
r	.66	.77	.94	.97	.95	.96	.96

Note. *M* = mean radius; *SD* = standard deviation; *CV* = coefficient of variance (%); *r* = coefficient of correlation.

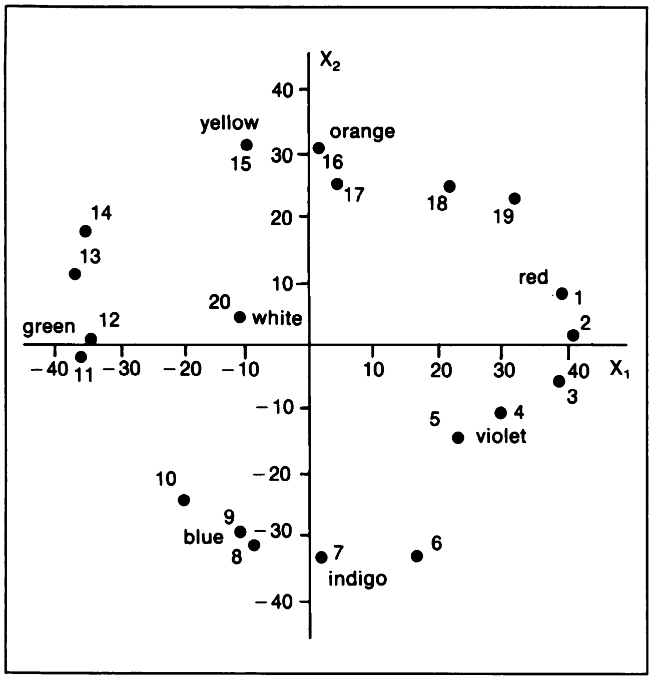

Fig. 3. Semantic space of artificial color names after repeated training and fixing. Subject V.L.

engaged in the semantic coding of information along two independent paths. The first path probably begins in the striate cortex, in the ending zone of the axons of relay cells of the lateral geniculate body. Here are formed the output signals of the color-opponent channels of the color analyzer that are responsible for the quick formation of the basic axes of the semantic space. The second path begins, presumably, in the poststriate areas of the sensory system, in the v2–v4 zones. From here originate the output signals of the color-detector cells responsible for the gradual shaping of the spherical structure of the color names in the semantic space, that is, the structure of color categories.

In conclusion, the semantic space of artificial color names representing color stimuli without other semantic links closely corresponds to sensory space. The similarity of the semantic space of artificial color names and the semantic space of natural color names supports the hypothesis that the dominant constituent in the semantics of natural color names is the sensory one. Our hypothesis is that the formation of color categories is a result of the activity of color-opponent channels and color-detector cells. The dynamics and the direction of the influence of these two factors on the shaping of the semantic color space differ substantially.

REFERENCES

Artemieva, Ye.Yu. (1968). *Subjective semantic psychology.* Unpublished doctoral thesis, Moscow State University, Moscow. (in Russian)

Fillenbaum, S., & Rapoport, A. (1971). *Structure in the subjective lexicon.* New York: Academic Press.

Izmailov, Ch.A. (1980). *Spherical model of color discrimination.* Moscow: Moscow State University. (in Russian)

Izmailov, Ch.A. (1982). Uniform color space and multidimensional scaling (MDS). In H.-G. Geissler, H.F.J.M. Buffart, E.L.J. Leeuwenberg, & V. Sarris (Eds.), *Psychophysical judgment and the process of perception* (pp. 52–62). Berlin: VEB Deutscher Verlag.

Izmailov, Ch.A., & Sokolov, E.N. (1991). Spherical model of color and brightness discrimination. *Psychological Science, 2,* 249–259.

Izmailov, Ch.A., Sokolov, E.N., & Chernorizov, A.M. (1989). *Psychophysiology of color vision.* Moscow: Moscow State University Publishers. (in Russian)

Miller, G.A. (1969). A psychophysical method to investigate verbal concepts. *Journal of Mathematical Psychology, 6,* 169–191.

Miller, G.A. (1971). Empirical methods in the study of semantics. In D.D. Steinberg & J.A. Jakobovits (Eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology* (pp. 569–585). Cambridge, England: Cambridge University Press.

Shmelev, A.G. (1983). *An introduction to experimental psychological semantics.* Moscow: Moscow State University. (in Russian)

Sokolov, E.N., & Vartanov, A.V. (1987). On the semantic color space. *Psikhologicheskii Zhurnal, 7,* 58–65. (in Russian)

Torgerson, W.S. (1958). *Theory and methods of scaling.* New York: Wiley.

(RECEIVED 5/29/91; ACCEPTED 10/21/91)