Color–concept associations reveal a new conceptual space

Kushin Mukherjee[a,1], Laurent Lessard[b], Michael Gleicher[c], Timothy T. Rogers[d,e,f], and Karen B. Schloss[d,e]

Affiliations are included on p. 11.

Edited by Martin S. Banks, University of California, Berkeley, CA; received September 23, 2025; accepted July 23, 2026

People are capable of using simple perceptual features such as color to signify complex meanings (perceptual semantics), an important aspect of nonverbal communication. This paper considers the nature and structure of the mental representations that support these abilities, using color semantics as a key example. Prior work has shown that people can judge the strength of association between essentially any color and any concept. Across four experiments we show that such associations express a low-rank conceptual space, color semantic space, whose dimensions can be well-characterized using color space regression. The color semantic space is distinct from classical connotative (affective) and denotative (literal) semantic spaces and from perceptual color space. Further, color semantic relations help to predict human judgments of word-meaning similarity across a diverse set of concepts—even when such judgments are *not* explicitly about color. The methodology we develop is general, offering a means of evaluating the mental structures underlying other aspects of nonverbal communication. Together, the work suggests that color–concept associations are encoded within a previously unrecognized conceptual representation space, and opens the possibility that human thought may rely on multiple different kinds of conceptual spaces—with implications both for cognitive science generally and for understanding the use of color in domains such as visual communication.

color–concept associations | conceptual similarity | low-dimensional psychological spaces | language models

A central issue in cognitive science concerns the nature, origins, and structure of the conceptual representations that allow people to communicate, understand, and reason about the world (1–4). Since Rosch's pioneering work (5, 6), a widespread view is that conceptual representations model the structure of the environment, as filtered through perception, action, and language (7–12). From this perspective, concepts can be viewed as points within a continuous *denotative* semantic space where items sharing many properties in the world lie near one another and are understood to be similar "kinds of things" (13). Thus wasp and butterfly are proximal because they share many properties and are both kinds of insect whereas *sun* is distal to both because it shares few properties with insects and is understood to be a quite different kind of thing. To estimate coordinates of concepts within a denotative space, researchers typically compute embeddings from human-generated datasets that rely on denotative meaning, such as semantic feature norms, semantic similarity judgments, or patterns of word-co-occurrence in natural language. The resulting *representational geometry* (i.e., patterns of interitem distances) captures both explicit judgments of meaning similarity (14, 15) and the tendency to generalize properties from one item to another during learning (8); for instance, learning that the butterfly "has plaxium inside" generalizes more to the (nearby) wasp than to the (distal) sun.

These ideas help to explain many aspects of behavior, but render others mysterious. Human knowledge, communication, and inference can hinge on mental associations that do not transparently reflect the structure of experience, suggesting that there exist conceptual spaces that go beyond a first-order statistical model of the environment. Indeed, cognitive science has generally recognized a second, qualitatively different connotative semantic space, revealed by early work on semantic differentials, showing that people can consistently rate concepts on anchored scales of polar adjectives even when these adjectives are not veridical attributes of the concepts (16, 17). For instance, people systematically judge whether tornadoes are "fair" or "unfair" or whether the sun is "dirty" or "clean." Across different cultures and languages, ratings on such scales exhibit low-dimensional structure with concepts falling along three latent affective dimensions: Evaluation (valence), Potency (strength), and Activity (energy), collectively known as EPA space (18–20). A concept's location within this space reflects its affective connotations rather than its literal denoted meaning: butterfly and sun lie near one another and far from

Significance

Human minds can infer complex meanings from quite simple percepts; for instance, colors in data visualizations may signify political allegiances, intensity of continuous variables, or categories of discrete data. Such inferences support our ability to communicate visually, but their cognitive underpinnings remain mysterious. Taking color semantics as a key example, we show that percept–concept associations are encoded within a previously unrecognized conceptual representation space distinct from classical connotative (affective) and denotative (literal) meaning spaces. We further show that such a *color semantic* space influences people's judgments about concept similarity even when the judgments are not about color. The work suggests that human cognition may encompass many different meaning spaces and provides a general methodology for investigating this possibility.

Author contributions: K.M., L.L., T.T.R., and K.B.S. designed research; M.G., T.T.R., and K.B.S. contributed new reagents/analytic tools; K.M. analyzed data; and K.M., L.L., T.T.R., and K.B.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: kushinm11@gmail.com.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2526862123/-/DCSupplemental.

Published September 16, 2026.

wasp because they evoke positive feelings whereas the wasp does not (21, 22). EPA space can also be recovered when embedding emotion words based on perceptual associations (colors, shapes, sounds, music) (23–25), and when embedding colors or music based on emotion associations (26–30). Thus, EPA space describes mental representations that connect words to perceptual features via affect, allowing people to relate otherwise incommensurate domains such as color and music [emotion mediation hypothesis (26–28)].

This paper considers whether other qualitatively different conceptual spaces may exist, taking color semantics as a key example. Work on visual communication has shown that people can judge the association strength between any color and any concept, and that such associations constrain the meanings people infer from colors in visualizations like graphs and maps (31–38). Moreover, such associations seem to contravene the representational geometries encoded by both denotative and connotative spaces: Concepts with similar denoted meanings and affective connotations can evoke dissimilar color associations (e.g. raspberries and blueberries), while items dissimilar in kind and affect can have similar color associations (e.g. night and socks). Thus, color–concept associations, and the inferences about color meaning they support, may be encoded within a color semantic conceptual space that aligns with neither denotative nor connotative spaces. If so, this raises the possibility that human thought is undergirded by multiple different conceptual spaces. Alternatively, color–concept associations may be mediated primarily by the EPA space, or by some combination of relations within denotative and connotative spaces; or the representations underlying such color–concept associations may not cohere as a conceptual space at all.

To adjudicate these possibilities, we operationalize criteria for a novel conceptual space as a form of mental representation that: 1) allows people to infer meanings of stimuli, 2) can be characterized as a continuous low-dimensional space, 3) supports generalization of information across items, and 4) encodes representational geometries that differ and are not predictable from other conceptual spaces.

Prior work has shown that patterns of color–concept association strongly influence the meanings people attribute to colors in graphical displays of information, for both abstract and concrete concepts and for concepts with both high and low color diagnosticity (32, 33, 36, 39, 40). Thus color–concept associations already meet criterion 1. We therefore assess whether they also meet the remaining criteria, and if so, whether human perceptions of word-meaning similarity encompass color semantic relations.

To evaluate whether color–concept associations express a low-dimensional representation space (criterion 2), Experiment 1 measured mean associations for 30 concepts across 71 colors spanning perceptual color space, then tested the resulting matrix for low-rank structure using singular-value decomposition. If patterns of color–concept association are largely arbitrary, a matrix encoding association strengths between n concepts and m colors would be nearly full-rank (i.e., would require the smaller of n or m orthogonal components to reconstruct). Alternatively, if patterns of color–concept association “hang together” in reliable ways, the matrix will be low-rank, that is, well-approximated by a small number of dimensions k where $k \ll n, m$.

To evaluate whether the resulting space supports generalization (criterion 3), we tested whether it predicts associations between colors and concepts not used to estimate the space. For colors, we first applied regression to understand how the dimensions of color semantic space vary with respect to six different metric properties of color. We then used the resulting models to predict the conceptual associations of new colors from their metric properties. For concepts, we evaluated whether the color semantic space predicts color associations for 40 new words belonging to different semantic categories. Significant above-chance prediction for both analyses provides evidence that color semantic space supports generalization.

We then evaluated whether the representational geometry of color semantics differs from that of connotative and denotative spaces (criterion 4). Experiment 3 used semantic differential ratings to estimate EPA (connotative) coordinates for words used in Experiments 1 and 2, while Experiment 4a estimated denotative coordinates via two natural-language processing (NLP) methods. In each case, we applied Procrustes correlation to quantify the alignment of representational geometries between color semantic space and alternative conceptual spaces. Finally, Experiment 4b used regression to assess whether color semantic information accounts for unique variance in the similarities people discern among word meanings.

Together, these experiments suggest that color–concept associations reveal a previously unrecognized conceptual space, which expresses representational geometries distinct from both denotative and connotative semantics. They also illustrate a general methodology for finding such spaces in other perceptual modalities, with implications both for our basic understanding of human conceptual knowledge and for practical applications, such as in visual communication.

Results and Discussion

Experiment 1: Evaluating Structure and Generalization of Color Semantic Space. This experiment investigated whether color semantic space a) is low rank and b) generalizes to held out colors and concepts. We tested these possibilities using color–concept association ratings data from human participants for each of 71 colors spanning perceptual color space [`UW-71` color library (36)] and each of 30 concepts spanning six broad semantic categories ranging in concreteness (Table 1). We structured the mean association ratings (averaged across participants) in a matrix X of 30 concepts by 71 colors (*Methods*).

Estimating the dimension of color semantic space. To determine whether color semantic space is low rank, we decomposed the

Table 1. Concepts used in Experiments 1 and 2 organized by category and the number of rater judgments used to estimate color–concept associations for each concept within each category

Category	Concepts	n
Experiment 1		
Clothes	Dress, Pants, Shirt, Socks, Shoes	48
Fruits	Blueberry, Lemon, Mango, Strawberry, Watermelon	46
Scenes	Beach, Field, Ocean, Sky, Sunset	45
Emotions	Angry, Disgust, Fearful, Happy, Sad	50
Directions	Above, Below, Beside, Near, Far	44
Times of Day	Dawn, Day, Dusk, Noon, Night	46
Experiment 2		
Animals	Bear, Bird, Lion, Frog, Fish	45
Automobiles	Airplane, Car, Boat, Truck, Train	51
Fruits	Apple, Banana, Cherry, Grape, Peach	46
Vegetables	Carrot, Celery, Corn, Eggplant, Mushroom	45
Weather	Blizzard, Drought, Hurricane, Lightning, Sandstorm	52
Values	Evil, Greed, Justice, Love, Peace	50
Activities	Driving, Eating, Sleeping, Leisure, Working	45
Properties	Comfort, Efficiency, Reliability, Safety, Speed	44

color–concept association ratings matrix X via singular value decomposition (SVD). In contrast to other matrix factorization techniques, SVD need not operate on a covariance matrix derived from a dataset, but can decompose X directly, even if asymmetric. To extract k orthogonal latent components from a matrix X with m rows (words) and n columns (colors), SVD models the data as the matrix product: $X_{m\times n} = U_{m\times k} \cdot D_{k\times k} \cdot (V^{\top})_{k\times n}$, where U encodes the direction each concept "points" in k dimensions, V^T encodes the direction of each color in k dimensions, and D is a $k\times k$ diagonal matrix whose entries indicate the singular values (importance) for each dimension, ordered from largest to smallest. The matrices $U \cdot D$ and $D \cdot V^T$ constitute embeddings of the words and the colors, respectively, within the same k-dimensional space. When the number of components k is equal to the smaller of m or n, the matrix product $U \cdot D \cdot V^T$ perfectly recovers the original matrix. If, however, the matrix is low-rank, it can be closely approximated by a much smaller number of components (see Fig. 1).

To test for low-rank structure, we reconstructed the original matrix using between 1 and 30 singular vectors and measured the quality of the reconstruction as the Pearson correlation between the original and reconstructed matrices. The blue curve in Fig. 1 (labeled "true data") shows that this metric increased from 1 to 6 components and leveled off when further components were added, suggesting that a **rank 6** approximation captures the association ratings well (average $r(69) = 0.96$; 94% of variance explained). The red curve (labeled "permuted mean") shows the average reconstruction correlation expected under the null hypothesis of no low-rank structure, computed via permutation sampling (*SI Appendix*); at rank 6 it was far lower than the observed correlation ($r(69) = 0.67$, 45% of variance explained) and 23 components were needed to reach the performance of 6 components for true data. Thus, the true color–concept association matrix expressed low-rank structure beyond that expected from a comparable random matrix.

The columns of matrix V^T in the decomposition encode how each of the 71 colors contributes to each latent dimension of the color semantic space, as shown in Fig. 2*A*, which plots the colors ranked from lowest to highest value along each dimension. To understand how a color's coordinate along each dimension relates to properties of its appearance, we fit a color-space regression model based on CIELCh space, previously established to capture patterns of variability in color preferences (41). The model includes predictors for lightness (L*) and chroma (C*), and captures variability in hue using the sine and cosine of hue angle (h) to obtain the "dominant hue" (first harmonic) and sin(2h) and cos(2h) to obtain the "dominant axis" (second harmonic). The first harmonic captures hue opponency (e.g., peak at red; trough at green) while the second captures weight on both ends of an opponent axis (e.g., peak at red and green; trough at blue and yellow). Using both sine and cosine as predictors allows us to represent any amplitude (strength) and phase (angle). Collectively, these predictors can be thought of as metric characteristics of colors that extend beyond the three-dimensional perceptual space to capture complex patterns, such as component 2 varying from light blues to darker reds and greens (Fig. 2*A*). The resulting regression models allowed us to both (a) interpret each dimension of color semantic space in terms of color properties and (b) predict conceptual associations for new colors not used to construct the space.

We fit a color space regression model to each of the 6 dimensions in Fig. 2*A*, with model predictions shown in Fig. 2*B*. The model predictions fit the components well ($r(69) = 0.55$ to 0.85, $Ps < 0.001$; Fig. 2 *A* and *B*), establishing a systematic relationship between the CIELCh color properties and each of the color semantic dimensions. Fig. 2*C* shows the regression weights for the dominant hue (*Top*) and dominant axis (*Middle*) for each dimension. Together, the dominant hue and dominant axis correspond to the more positive values in Fig. 2 *A* and *B*, and the opposing hue/axis correspond to the more negative values. The dominant hue plots (Fig. 2*C*) show hue angle in polar coordinates, with the dominant hue shown as a black dot and its opposing hue as a white dot. The radius represents the weight of the dominant hue in the regression model, corresponding to its strength. For example, component 1 has a relatively weak dominant hue around purplish-blue and component 2 has a stronger dominant hue around blue. The dominant axis plots are similar (Fig. 2*C*), but show the dominant axis of opposing hues as a pair of black dots and the opposite axis (rotated 90 degrees) as a pair of white dots, with radius indicating weight in the regression model. For example, component 1 has a relatively weak dominant axis from purplish-blue to greenish-yellow and component 2 has a stronger dominant axis from greenish-blue to orangish-yellow. Fig. 2*C* (*Bottom*) shows the weights on lightness and chroma, with black (white) dots indicating positive (negative) weight. For example, component 1 has a strong positive weight on darker, lower chroma colors and component 2 has a less strong positive weight on lighter, lower chroma colors. The combination of the dominant hue, dominant axis, lightness, and chroma gives rise to the patterns in Fig. 2*A*, as illustrated by regression equation

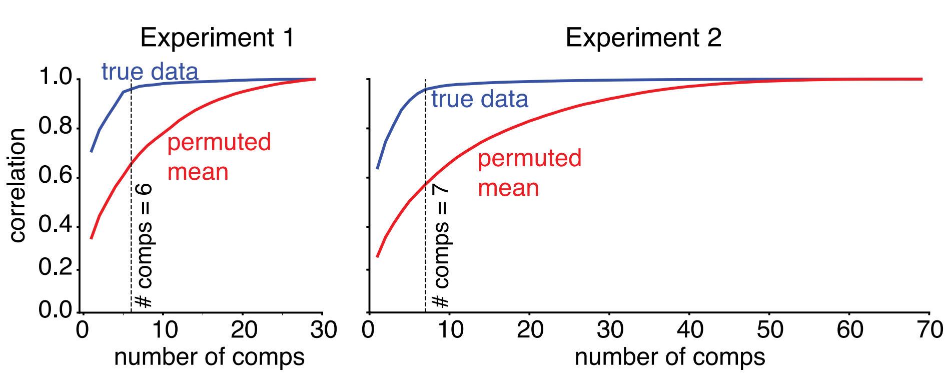

Fig. 1. Pearson correlation between the data matrix X of concepts × colors and its SVD reconstruction $U \cdot D \cdot V^T$ as the number of components varies (blue curve; "true data"). For Experiment 1, using 6 components for 30 concepts achieves a mean $r = 0.96$. A corresponding analysis for randomly shuffled data (red curve; "permuted mean") shows the low-rank structure is lost. Experiment 2 shows a similar pattern, where using 7 components for 70 concepts achieves a mean $r = 0.96$.

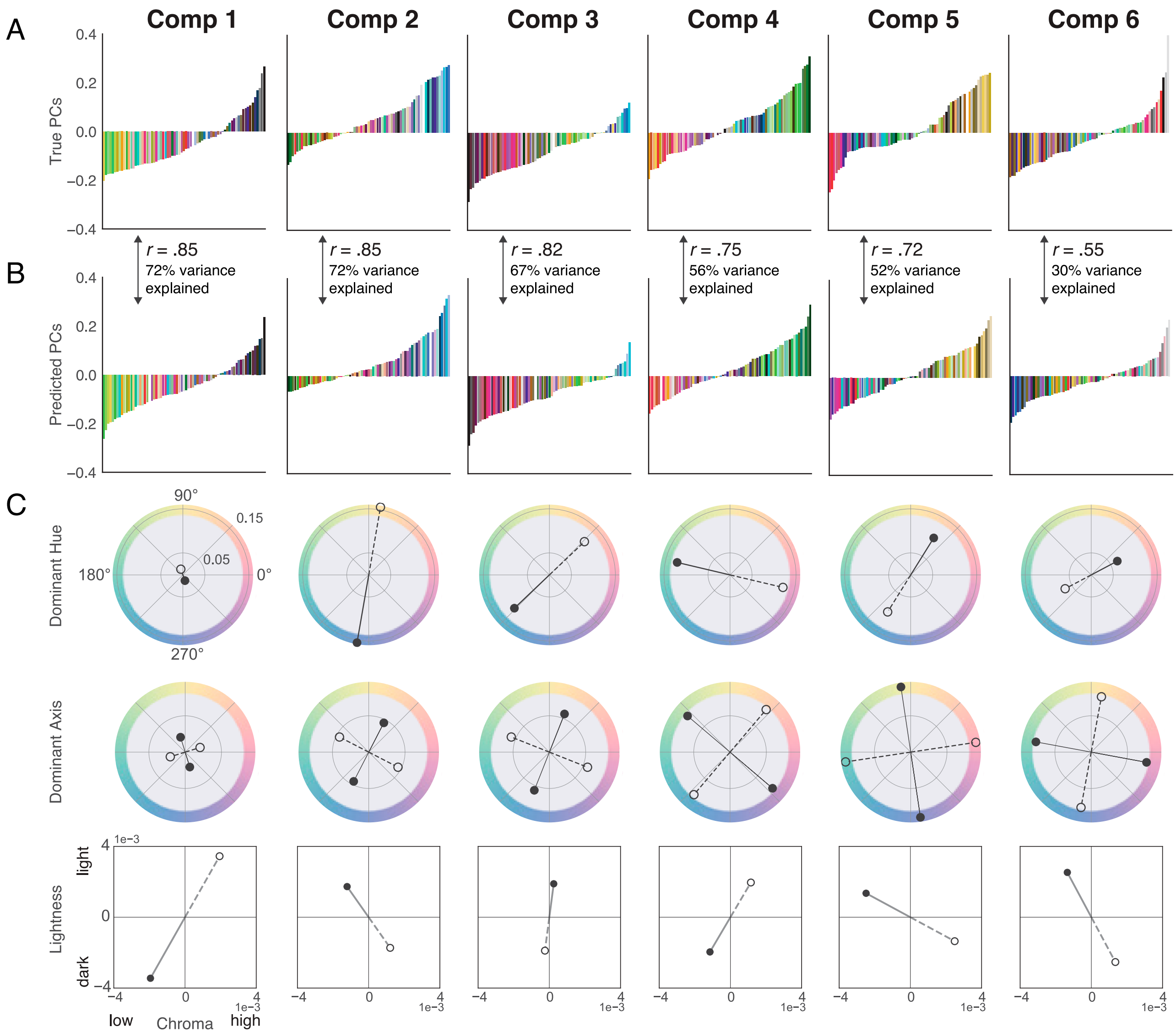

Fig. 2. (*A*) Loadings for each UW-71 color on each of the 6 color semantic space dimensions estimated via SVD sorted from lowest to highest. (*B*) Predicted loadings on each color using color-space regression models. Arrows between (*A*) and (*B*) indicate Pearson correlations between true and predicted loadings. (*C*) Dominant hue (first row), dominant axis (second row), and lightness and chroma (third row) for each of the six components. Peaks are black circles and troughs are white circles. See text for details.

predictions in Fig. 2*B* (See *SI Appendix*, section 3 for further details).

Together these results show that the latent space underlying color–concept associations meets two additional criteria for a conceptual space: It is low-dimensional and, by virtue of its systematic relation to measurable properties of colors, might support generalization of information about conceptual associations to new, held-out colors (i.e., colors not used to estimate the space).

We further note that this color semantic space differs qualitatively from perceptual color space. Capturing the latent structure of the color–concept association matrix requires minimally 6 dimensions, complexity that cannot be captured by the three-dimensional nature of perceptual color similarity (for those with typical trichromatic vision). To see this, consider that a given item can have strong associations with multiple perceptually distinct colors; for instance, watermelons are strongly associated with both red and green (i.e., opposites in a perceptual color space), but not blue or yellow. Capturing such patterns requires that perceptually distal colors contribute similarly to a given color semantic dimension—for instance, reds and greens contribute similar, negative weights to the second component shown in Fig. 2*A*. Because this is so, relating metric properties of color to color semantic coordinates requires predictors that contort perceptual space, such as the dominant axis information shown in Fig. 2*C*.

Generalization to new colors. The preceding analysis suggests we should be able to predict associations of a new color (i.e., one not used to construct the space) with all 30 concepts by right-multiplying its CIELCh coordinates by the matrix product $U \cdot D \cdot W$ as shown in Fig. 3*B*. We tested this form of generalization using leave-one-out cross-validation: Given the color–concept association matrix of 30 concepts and the `UW-71` colors, X, we iteratively held out one color (column) and decomposed the resulting 30×70 matrix of associations as in Fig. 3*A*. We then predicted the held-out color's association with each concept as in Fig. 3*B* and computed the correlation between this prediction and the color's true pattern of concept associations (i.e., the held-out column of X). We repeated this process for each color. The resulting correlations (Fig. 4*A*) were significantly positive for all colors after applying the Holm–Bonferroni correction (mean $r(28) = 0.87$, SD $= 0.08$, range: 0.51 to 0.98, $P < 0.001$

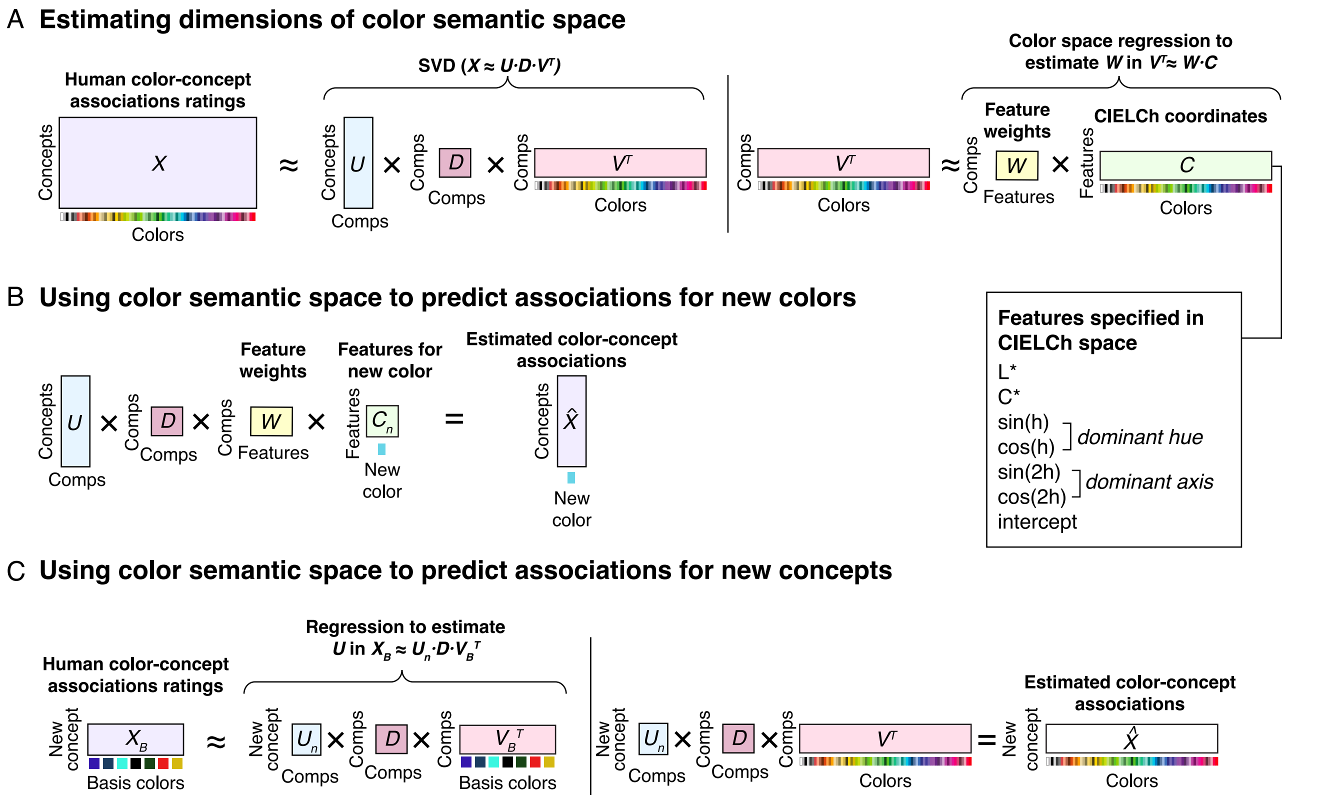

Fig. 3. (*A*) Our process for estimating color semantic space from human color–concept association ratings. The matrix of color–concept associations (X) is decomposed using singular value decomposition (SVD) into matrices U, D, and V^T. Matrix V^T is further decomposed using regression into a matrix of feature weights (W) times colorimetric feature coordinates (C). Each row of matrix V^T constitutes one of the dimensions of color semantic space. (*B*) Method for estimating the associations for a new color with all concepts using the color semantic space and colorimetric predictors. (*C*) Method for estimating associations for a new concept with all colors.

for all). Thus, color semantic space, characterized via color space regression, supports generalization to colors not included in the original dataset.

Generalization to new concepts. The low-rank structure of color semantic space further suggests we should be able to predict the full complement of color associations for a new concept (i.e., one not used to construct the space) from its association ratings with a small, carefully chosen set of basis colors—colors selected to provide the most information about a concept's loading on the 6 color semantic dimensions. Using the Joshi and Boyd method (42), we found that 7 such "basis colors" were sufficient to capture the full color semantic space: dark purple, dark blue, light blue, black, dark green, red, and yellow (*SI Appendix*, Fig. S2). Given color–concept association ratings between the new concept and these basis colors, we used linear regression to estimate the concept's directions U, then multiplied the result by the matrix of color directions V^T to obtain predictions for the concept's associations with all other colors (Fig. 3*C*).

We tested this form of generalization using leave-one-out cross-validation: From X we iteratively held out one concept (row) and decomposed the resulting 29 × 71 matrix of associations (Fig. 3*A*), then predicted the held-out concept's association with each color using only its associations with the 7 basis colors (Fig. 3*C*). We computed the correlation across the remaining 64 colors between predicted and true association strengths (i.e., held-out row of X), repeating this process for each concept. The resulting correlations (Fig. 4*B*) were significantly positive for all but one concept ("beside") after correcting for multiple comparisons (mean $r(62) = 0.81$, SD $= 0.17$, range: 0.20 to 0.97, $P < 0.001$ for all but one). Thus, color semantic space supports generalization to concepts not used to estimate the space: Association data from just seven basis colors were sufficient to estimate associations between the new concepts and any color.

In summary, Experiment 1 suggests that color semantic space meets essential criteria for a conceptual space: It is low-rank (i.e., low-dimension, criterion 2) and supports generalization of information about color–concept associations both for new colors (via their color space coordinates) and new concepts (via their associations with a few colors; criterion 3). Moreover, each dimension of color semantic space can be interpreted as encoding a different linear combination of 6 metric properties of color captured by CIELCh color space.

Experiment 2: Assessing Generalization of Color Semantic Space to Concepts Drawn from New Categories. Experiment 2 investigated whether the color semantic space from Experiment 1 generalizes to concepts belonging to semantic categories not used to construct the space. In Experiment 1, generalization to the new concepts may simply reflect shared category membership of the held out concept with the concepts that were not held out, with each category possessing its "own" color semantic space. If so, the model should fail to generalize beyond the original six categories. Successful generalization to new categories would instead suggest a common, category-independent color semantic space.

To test this idea, we evaluated how well the space constructed from the 30 concepts in Experiment 1 predicts associations for 5 new items from each of 7 new categories—animals, vegetables, vehicles, weather phenomena, human values, activities, other abstract properties—plus 5 additional fruits with distinct color associations from Experiment 1 (Table 1; see *Methods*). If strong

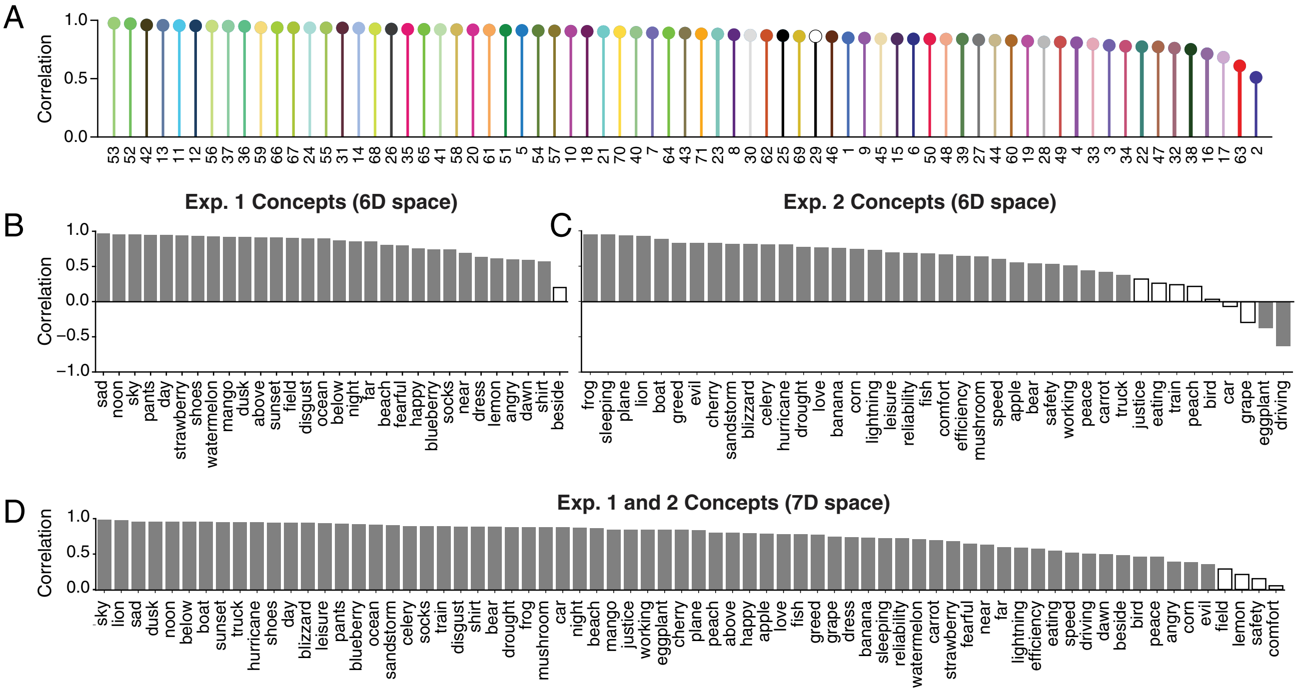

Fig. 4. (*A*) Pearson correlations between predicted and true association ratings for each UW-71 color with all concepts Numbers on the x-axis correspond to the index of the color in the UW-71 colors (see Table S1 in Supporting Information). (*B*) Correlations between predicted and true association ratings for each concept with all UW-71 colors, from Experiment 1. (*C* and *D*) Similar to (*B*), but for Experiment 2, using the color semantic spaces from Experiments 1 and 2, respectively. Filled-in bars are statistically significant at the 0.05 level after correcting for multiple comparisons. All correlations are computed in leave-one-out fashion.

generalization in Experiment 1 was due to test items sharing categories with training items, the space should generalize better to new fruits than to concepts in other new categories (e.g., animals, automobiles). Alternatively, if the space extends across domains, it should generalize similarly well to the new fruits and concepts from new categories.

Using the mean color–concept association ratings for the 40 new concepts and `UW-71` colors (*Methods*), we followed the procedure in Fig. 3*C* to 1) situate each concept in the color semantic space from Experiment 1 using association data from the 7 basis colors, 2) estimate that concept's associations with the remaining 64 of the UW-71 colors, and 3) correlate these estimates with participants' ratings to assess performance. Predicted associations correlated significantly and positively with true associations for 33 of the 40 novel concepts (Fig. 4*C*). Among the items showing nonsignificant or negative correlations, 2 were from the fruit category, whereas the concepts with the 5 highest correlations contained no fruits. These results suggest that good generalization does not arise only for concepts drawn from categories used to construct the semantic space.

Nevertheless, the mean predicted–true correlation across the 40 concepts in Experiment 2 (mean $r(62) = 0.54$, $SD = 0.38$; see Fig. 4*C*) was significantly lower than the mean correlation for the 30 concepts in Experiment 1 ($r(62) = 0.81$) (Fig. 4*B*). This result suggests that the original six dimensions from Experiment 1 were not sufficient to fully capture color semantic structure among the new concepts. Indeed, the rank 6 solution observed in Experiment 1 may have arisen precisely because we drew concepts from 6 semantic categories (Table 1). If so, then the number of components needed to model the expanded set of associations in Experiment 2 should scale approximately linearly—adding 7 semantic categories to the original 6 should roughly double the required number of components. Conversely, if the color semantic space from Experiment 1 is domain-general, then adding far fewer components should suffice to capture color–concept associations for all 70 items (combining those from Experiments 1 and 2).

To adjudicate these possibilities we aggregated data from both experiments (70 concepts) and repeated the procedure in Fig. 3*A*, computing correlations between true and reconstructed ratings using rank 1 to 70 matrices (Fig. 1, *Right*). Reconstructing the ratings matrix with comparable performance to the 6 components in Experiment 1 ($r(69) = 0.96$; 93% cumulative variance) required 7 components, far fewer than the dozen that would be required if rank scaled linearly with the number of semantic categories. Descriptions of each of the 7 components in terms of color space regression models can be seen in *SI Appendix*, Fig. S3. These results provide further evidence that color semantic space is indeed low rank and scales sublinearly with the set of concepts used to construct the space.

To assess whether the addition of a seventh latent component produced better generalization to the 40 novel items than did the original 6D space, we used the 7D space to estimate association ratings for each of the 40 concepts from Experiment 2, applying the same leave-one-out procedure from Experiment 1 (Fig. 3*C*). In this case the Joshi and Boyd method suggested that 11 basis colors were optimal for estimating a held out concept's 7D loadings. Using color–concept association ratings for the 40 concepts on those 11 colors, we generated predicted associations on the remaining 60 colors. The correlations between predicted and true values were significantly positive for all but 4 concepts, and the mean correlation of $r(58) = 0.73$ was significantly higher than the mean correlation for predictions from the original 6D space, $r(62) = 0.54$ ($t(39) = 3.17$, $P < 0.01$; Fig. 4*D*). All Pearson r values were first transformed to normally distributed values using Fisher's r-to-z transformation prior to comparisons using t tests. We also compared the predicted vs. true association strengths for just the original 30 concepts from Experiment 1 using the 7D space ($r(58) = 0.77$) vs. the 6D space ($r(62) = 0.81$) and found they did not significantly differ ($t(29) = -0.57$, $P = 0.57$). Together these results suggest that the addition of the seventh dimension did not further improve estimates for the original 30 concepts, but that it was needed to represent the additional 40 concepts introduced in Experiment 2.

In summary, Experiment 2 revealed that color semantic space a) remains low-rank when items from many new semantic categories are added to the estimation set (criterion 2), requiring only one additional dimension to capture an additional 40 items; and b) generalizes well to items from new semantic categories (criterion 3). Predicted vs. true association ratings for each color–concept pair are in *SI Appendix*, Figs. S6–S10 spanning Experiments 1 and 2.

Experiment 3: Comparing Color Semantic Space to EPA Space. Results to this point establish that color semantic representations meet the criteria for a conceptual space noted in the introduction. A remaining question is whether this space is distinct from other known conceptual spaces (criterion 4). For instance, it may be that color–concept associations are mediated by the same affective representations captured by connotative semantics (43). If so, the coordinates of words in the color semantic space should be highly correlated with the coordinates of the same words in the connotative space. Alternatively, color–concept associations may be mediated by a qualitatively different representational structure. In that case coordinates of words within color semantic space should *not* strongly correlate with coordinates in EPA space. Adjudicating these hypotheses requires measuring the alignment between two spaces, each encoding coordinates for the same set of items—in this case, embeddings of words.

To this end we applied Procrustes alignment (44), which finds a rigid transformation (translation, scaling, rotation, and reflection) of items in a *source* space that matches each point as closely as possible with the corresponding point in a second *target* space. Alignment is measured by the Procrustes correlation between aligned spaces, a multidimensional analog of Pearson's r (*SI Appendix*). Since Procrustes alignment finds the best possible match between spaces, it yields positive correlation values even for points randomly situated in each space—thus the significance of a positive correlation is evaluated with respect to a null distribution simulated via permutation sampling. The global representational geometries of two spaces are significantly related if their measured Procrustes correlation lies above the significance cutoff in the null distribution—if not, the geometries of the two spaces are no more similar than expected by chance.

Experiment 3 applied the method of semantic differentials to estimate EPA space coordinates for the 70 concepts from Experiments 1 and 2, then measured the Procrustes correlation between EPA and color semantic coordinates. Specifically, we collected new ratings for each concept on Osgood et al.'s 54 semantic differential scales (16, 18) (e.g., rate the concept "apple" on a scale from dry to wet). See *SI Appendix* for details and *SI Appendix*, Table S7 for the full set of scales. We then organized the mean ratings into a concept × scale matrix and estimated EPA coordinates for each of the concepts following the same dimensionality reduction techniques used in prior work (16, 18).

The first three components collectively accounted for 76.44% of the variance in the rating matrix, and the top 6 scales loading on each component were similar to the scales that were originally identified as highly loading on each component by Osgood (18), suggesting good replication of the EPA space for our 70 concepts (*SI Appendix*, Table S8).

To match the three dimensions of EPA space, we estimated a 3D embedding of the words based on their color–concept associations. We then computed the Procrustes correlation (44) of the 70 concepts in the two spaces (*SI Appendix*, Fig. S1*A*). While we observed a moderate global correlation between the two spaces ($r(68) = 0.30$, $P < 0.01$), the individual dimensions of the two spaces were not all aligned whether assessed using Procrustes alignment or more stringent regression methods. *SI Appendix*, section 5 outlines these other approaches including Procrustes alignment in 7 dimensions. Thus, given that EPA space only explains approximately 9% of the variance in color semantic space and the individual dimensions do not reliably align, our findings suggest that color semantic space and EPA space are not reliably similar for a general set of concepts. While prior work had shown that EPA space could account for people's associations between colors and emotion words (23), our results suggest EPA cannot account for associations between colors and a broader set of concepts, including concepts with similar affective connotation but highly different colors (e.g., blueberry and peach), see also *General Discussion*.

Experiment 4: Comparing Color Semantic Space to Denotative Spaces and Human Judgments of Meaning. Finally, we consider whether color semantic space is distinct from denotative semantic spaces. Further, since denotative semantics are often construed as the "default" meanings people discern, we additionally consider whether human judgments of word-meaning similarity ever reflect structure unique to color semantics. Because the data needed to estimate human-perceived similarities grows exponentially with the number of items (*Methods*), Experiment 4 employed the 30 words used in Experiment 1.

Experiment 4a: Alignment of color semantic and denotative semantic spaces. We adopted two approaches to estimation of denotative semantic spaces, each broadly representative of modern techniques: word2vec (45) trained on the COCA-fiction corpus (46), whose embeddings reflect co-occurrence statistics of words in large-scale language corpora; and word-embeddings extracted from GPT-3.5, a large language model (LLM) (47, 48) trained on web-scale corpora of text-based language data. We used GPT-3.5 as it was the most performant model that allowed extraction of single word embeddings via its API.

Fig. 5 shows hierarchical cluster plots of word-to-word similarities in color semantic and both denotative spaces. Both denotative spaces largely capture semantic category information, for instance grouping spatial terms together and separating these from emotions and times of day. Color semantic space splits such categories across clusters, expressing more abstract relations—for instance, grouping some spatial relations (below, far) with negative emotions (angry, sad) with later times of day (night, dusk).

To quantify alignment across spaces, we again reduced embeddings in both spaces to three dimensions via multidimensional scaling, then computed Procrustes correlations between each pair of spaces (r values in Fig. 5). The two denotative spaces correlated significantly with each other ($r = 0.76$, $P < 0.001$), showing that the different methods lead to similar representational geometries. Neither denotative space correlated significantly with color semantic space (word2vec: $r = 0.33$, $P = 0.14$; GPT-3.5: $r = 0.34$, $P = 0.11$). *SI Appendix* reports additional approaches to measuring alignment, all suggesting no relationship between

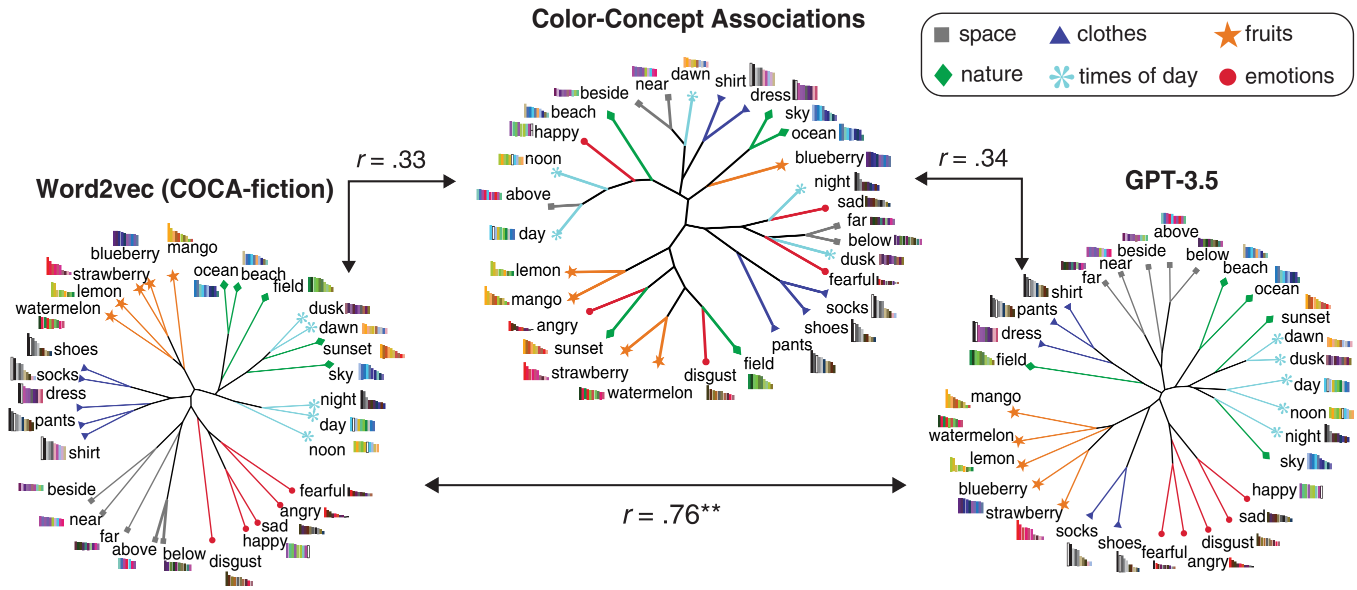

Fig. 5. Hierarchical cluster dendrograms of the 3D multidimensional scaling coordinates for (*Top*) color–concept association ratings, (*Left*) word2vec embeddings, and (*Right*) GPT-3.5 embeddings for each of the 30 concepts from Experiment 1. The colored bars next to each concept indicate the most associated colors with that concept from the UW-71 colors. Arrows between dendrograms indicate Procrustes correlation values. The color–concept association dendrogram displays an organization of concepts based on similarity of color association as opposed to semantic category. This organization is indicated by terminal leaves on many branches having different shapes corresponding to different categories. GPT-3.5 and word2vec primarily organize concepts based on semantic similarity.

denotative and color semantic structure. Thus, color semantic space is distinct from both connotative and denotative spaces, meeting criterion 4 for a novel conceptual space.

Experiment 4b: Human judgments of word meaning reflect color semantic structure. To understand whether color semantic similarities influence human perception of word-meaning, we used a triadic comparisons task to estimate word-embeddings directly from human judgments of semantic similarity in different task conditions. On each trial, participants decided which of two option words was most similar to a target word in some respect. From many such judgments one can embed the corresponding words in a low-dimensional space such that items nearby in the space are often selected as similar to one another (49–53).

Typically when participants judge similarity without any further instruction, the resulting word embeddings express a representational geometry that correlates with denotative semantics (53–55). When instructed to focus on a particular semantic dimension such as size or category, however, geometries of the resulting embedding change to emphasize the information to be retrieved (56). The task thus provides a means of assessing varieties of semantic structure discerned in different task contexts. We collected data in three task conditions, each testing a different hypothesis about when similarity judgments will reflect color semantic structure. First, participants may only access color semantic information when explicitly thinking about color–concept associations; thus the color-only condition asked them to judge based solely on colors associated with the words. Second, participants may discern color semantic relations in any task requiring abstract, metaphorical thought; thus the symbolic condition asked them to judge which option was best used as a symbol or metaphor for the target. Third, color semantic relations may influence perceived meaning similarity more generally; thus the kind task asked participants to judge which option referred to a more similar "kind of thing."

From the resulting data we computed a 3D word embedding separately in each condition, yielding three different representational geometries. We then fit regression models to predict each word's coordinates along each dimension within each space from its coordinates in the 3D denotative spaces (running separate analyses for the COCA and GPT-3.5 embeddings) and the first 3 dimensions of color semantic space. This effectively removed variation in the human-derived space explained by denotative semantic relations. If color semantic relations explain unique variance in perceived word meaning in a given task condition, coordinates in color semantic space should then account for significant additional variation in the residuals.

As expected from prior work, denotative predictors from both GPT-3.5 and word2vec explained significant variance under all three conditions, but critically, color semantic relations explained additional unique variation in all three dimensions of the human-derived embeddings for all three task conditions (Fig. 6*B*). Color semantic coordinates explained the most variation for color similarity judgments, but also explained significant variation in both the symbolic and kind-based judgments–strongly suggesting that color semantics influences perceived similarity, not just when color is directly relevant or in contexts requiring abstract thought, but also in more general contexts invoking word meanings.

General Discussion

Human concepts encompass knowledge of highly abstract relationships that go beyond the immediate sensory-motor structure of the environment (8, 57–59). While early work on associations between percepts, like color, and words (concepts) focused on affect ("connotative semantics") as the critical medium for representing such knowledge (17, 18), the current work reveals a qualitatively different kind of abstract, cross-modal conceptual structure: a color semantic space. People can report association strengths between essentially any color and any concept (34); we have shown that these associations encode a latent space possessing key hallmarks of other conceptual structures: 1) they allow people to infer meanings of percepts (32, 33, 39),

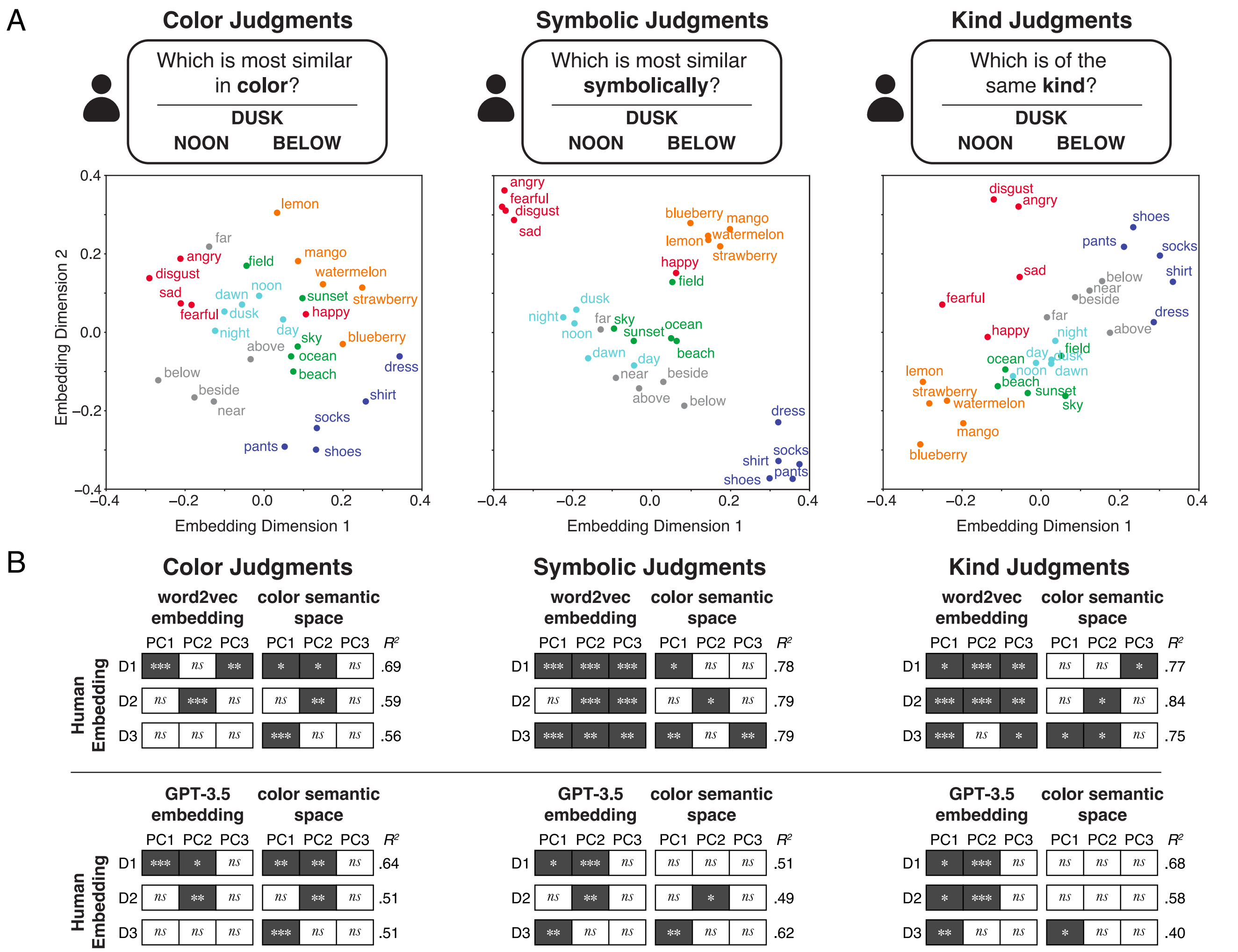

Fig. 6. (*A*) Structure of triplet behavioral experiment and visualizations of the concepts from each condition (color, symbolic, kind) using the first two embedding coordinates. (*B*) Regression coefficient *t*-values for models predicting the 3-dimensional triplet embedding coordinates from word-embedding and color–concept association components. Asterisks indicate coefficients that reliably improve model fit with $P < 0.05$,** $P < 0.01$,*** $P < 0.001$. Empty boxes indicate coefficients that were not statistically significant. R^2 values next to each row correspond to variance explained in each human embedding dimension by language embedding and color–concept association components.

2) they exhibit low-rank structure that differs from perceptual color space, 3) they support generalization of color–concept associations across all of color space and from a small set of "basis" colors to a much broader set of linguistic concepts (Experiments 1 and 2). We have also shown that the space is novel insofar as it encodes representational geometries among words distinct from those encoded in both the connotative semantic space revealed by Osgood and colleagues (Experiment 3) and from denotative spaces estimated from contemporary natural language processing techniques (Experiment 4a). Moreover, color semantic relations explain unique variance in human judgments of meaning similarity in tasks explicitly requiring color association, tasks demanding symbolic/metaphorical thinking, and even tasks tapping semantic knowledge more broadly (Experiment 4b)—suggesting that color semantics permeates the conceptual relations people discern in word meanings generally.

Understanding how the human mind represents associations between concepts and colors is fundamental for understanding the human ability to interpret colors as "standing for" a variety of ontologically distinct meanings, from temperatures to emotions to political identities to sports teams (60–62). Color–concept associations constrain the meanings people derive from colors in graphical presentations of data (e.g., bar graphs) (39, 63–65) and determine when people can discriminate between the meanings of different colors in a graph (32, 33, 36). The present work shows that these abilities rest on a low-dimensional psychological space (color semantic space) that is distinct from denotative and connotative spaces. We use the term color-semantic to describe the space our work uncovers because color–concept associations critically constrain the meanings that people attribute to colors in visual communication. By this we do not intend to imply that color-semantic knowledge is independent of color perception, nor that it relies solely or predominantly on linguistic semantics. Our focus is not on the question of whether color-semantics is mainly perceptual or mainly conceptual, but rather on the question of how color-semantic structure relates to the cross-modal sensory, motor, and linguistic structure of the environment. The latter is an important topic that we hope future work will address.

The current results differ from established literature in two important respects. First, most prior work on the dimensional structure of color–concept associations focused on emotion concepts, with EPA space emerging as the dominant organizing framework (16). Here we showed that emotions are well situated in color semantic space alongside both highly abstract concepts (properties, values) and highly concrete ones (fruits, vegetables). This implies a) theories of color–concept associations need not treat emotions separately, and b) prior approaches to color–emotion associations via affective/connotative semantics do not scale to the wider concept range studied here, since their latent structure cannot be recovered from purely connotative semantic structure.

Second, little prior work has considered whether color semantic knowledge impacts word meanings outside color-meaning interpretation contexts. Experiment 4b suggests it does—color semantics explains unique variance in judgments about word meanings when color is not central to the task. Future work can investigate to what extent other sources of perceptual knowledge—tactile texture, sound, olfaction—might also influence semantic judgments. As mentioned at the outset, the present approach taken for color semantics can be extended to these sensory domains to potentially uncover modal semantic spaces, which might each mediate percept–concept associations. Characterizing such modal spaces might be critical for developing generalized theories of how percepts and concepts are represented in the mind.

Our results raise the important question of the role of language in acquiring color semantics. Prior work suggested color–concept associations may be learned from language's higher-order covariance structure. For example, Liu et al. (66) showed that distributional language models (e.g., GPT) predict color–concept associations comparably for sighted and blind participants. Moreover, Kim et al. (67) argued that congenitally blind participants' linguistically transmitted causal knowledge (e.g., photosynthesis makes plants green) produces shared associations with sighted individuals. However, our results challenge this: Color semantic relations could not be recovered from embeddings derived from word2vec or a high-performance language transformer (GPT 3.5; Experiment 4). An alternative is that such knowledge arises from cross-modal experiential structure—not just the structure of language alone, but how language connects to visual and sensory-motor experience. Indeed, this proposal accords with contemporary thinking about semantic knowledge acquisition generally (68).

Related open questions concern why color semantic dimensions are structured as they are. For instance, while component 1 in our present data distinguishes blues from greens and darker/low-chroma from lighter/high-chroma colors, the environmental characteristics that produce these patterns remain unclear. Targeted experiments exposing participants to novel concepts under varying environmental color statistics might reveal how environments shape people's color semantic space. For instance, Schoenlein et al. (69) taught participants novel words for visual objects ("aliens") varying in shape and color. Each label denoted aliens with a distribution of possible colors, with some colors more likely than others, though shape alone could determine the correct label. After learning, participants associated newly learned words more strongly with colors more likely for the corresponding alien shape—consistent with associations arising from cross-modal sensory learning. Ongoing work is assessing whether higher-order cross-modal structure can support learning more abstract word-to-color associations.

Finally, while we have emphasized how our results advance our understanding of human conceptual representation, the current work also has important practical implications when it comes to human visual communication. Color is a core visual channel through which information is transmitted in information visualizations (35, 70, 71). When color is used to convey meaning, color semantic structure strongly influences the ease with which intended meanings are communicated (32–34, 36, 37, 39). Our work suggests that such associations can be efficiently computed for essentially any concept or color with only a handful of human measurements, so long as the measurements are selected so as to fully span the color semantic space. This opens up the possibility for scalable design of bespoke visualizations, whose effectiveness rests on metrics derived from such human color–concept associations.

Methods

Color–Concept Association Rating Task. Experiments 1 and 2 used a rating task to assess people's associations between each color from the `UW-71` color library and concepts that spanned a variety of categories (e.g., fruits, emotions, scenery; see Table 1).

Participants. Participants were UW-Madison undergraduates who received extra course credit for their participation. All participants in this and subsequent experiments gave informed consent, and the UW-Madison IRB approved the experimental protocol. Experiment 1 tested 315 participants (M age $= 20.67$, with gender self-reported as 174 women, 139 men, 2 other). We excluded $n = 36$ for atypical color vision, assessed using 1) two self-report questions asking if the participant had difficulty distinguishing between colors relative to the average person and if they considered themselves colorblind (excluded if they reported "yes" to at least one question) and 2) using digital Ishihara plates (excluded if incorrectly identified the number in more than 1/6 Ishihara plates); final $n = 279$. In Experiment 2, data for fruits, vegetables, properties, and activities were collected as part of Mukherjee et al. (36) ($n = 185$; M age $= 18.66$, 99 women, 86 men; 5 excluded for atypical color-vision, final $n = 180$). Data for the remaining concept categories (Table 1) were collected in a second batch ($n = 220$; M age $= 18.59$, 142 women, 77 men, 1 nonbinary; 22 excluded for atypical color-vision, final $n = 198$).

Design, displays, and procedure. Participants were randomly assigned to one concept category (e.g. fruits, emotions, scenery) and rated associations between each of 71 colors (*SI Appendix*, Table S2) and five concepts within that category (Table 1). The UW-71 color library includes 71 colors, with 58 sampled uniformly in CIEL*a*b* space ($\Delta E = 25$) and an additional 13 colors at L $= 88$ to accommodate a greater range of greens and yellows (36).

Experiment 1 employed 30 concepts spanning 6 different conceptual categories (clothes, directions, fruits, times of day, scenes, and emotions) varying in abstractness and the specificity of their color–concept associations. Experiment 2 added 7 additional categories also varying in abstractness and specificity (see Table 1 for the full set). Both used the same blocked randomized design and procedure, with participants rating associations between each of the 71 colors for a given concept (in a random order) before going on to the next. Concept order was randomized. On each trial, a single `UW-71` color was shown as a color patch (80 px $\times$ 80 px) centered on the screen with the concept word displayed above (font-size: 24 pt, font-family: Lato). Below the colored square was a line-mark slider scale (400 px long), with the left end labeled "not at all" and the right end labeled "very much" and the center marked with a vertical line (3 px wide and 32 px tall). Participants rated how much they associated the given color and concept by moving the slider marker and clicking a "continue" button to record their response. Trials were separated by a 1000ms intertrial interval. The background was gray (CIE Illuminant D65, x $= 0.3127$, y $= 0.3290$, Y $=$ 10 cd/m^2) so that very dark colors (e.g., black) and very light colors (e.g., white) could be seen against the background. Participants completed the experiment online on their own devices and the color coordinates were translated from CIELAB to RGB values assuming a standard sRGB display. The precise colors that each participant saw varied with the specifications of their monitor as is typical in online color experiments, which aim to be generalizable across displays (36, 37, 72–74). The experiment was coded in jsPsych (75, 76) for online data collection.

Before beginning the task, participants completed an anchoring procedure so they knew what associating "not at all" and "very much" meant to them in the context of these colors and concepts (77). They were shown a screen with all 71 colors and the five concepts they had been assigned. They were asked to identify the color they most and least associated with each concept.

Semantic Differentials Task. Experiment 3 assessed the 70 concepts from Experiments 1 and 2 on Osgood's semantic differentials rating task (18) using the same scales.

Participants. We tested 361 UW-Madison undergraduates who received extra credit in their Psychology course.

Design, displays, and procedure. Data were collected in two phases. The first phase focused on the 18 scales that loaded most heavily on the first three components in Osgood et al. (18). The full set of scales and number of raters can be seen in *SI Appendix*, Table S7. To reduce the number of ratings for each participant, we divided the 18 scales into 3 sets of 6 scales and randomly assigned participants to sets. Within sets, we also randomly assigned participants to judge either the 30 concepts from Experiment 1 or the 40 concepts from Experiment 2. We used a blocked randomized design, such that participants rated all of their assigned concepts on a given scale (random order) before going on to the next scale (scale order was also randomized). On each trial, participants saw a concept word in the center of the screen with a ratings slider below it. The ends of the slider were labeled with the two endpoints of the scale they were being asked to rate the concept on. Participants could use their cursor and click to indicate their associations for the concept on the scale.

The second phase tested the remaining scales from refs. 16 and 18 comprising a total of 54 items sans the 18 items collected in the first phase. We divided the remaining 36 scales into 6 sets of 6 scales and once again collected ratings for all 70 concepts on these new scales as described for the first phase. We aimed to collect $\sim$20 ratings per concept-scale combination.

Triadic Comparisons Task. Experiment 4b used triadic comparisons, implemented using the SALMON online tool (78), to estimate embeddings that capture human-perceived semantic similarities among the 30 words from Experiment 1 under different task instructions (49).

Participants. 36 UW-Madison undergraduates completed the experiment in exchange for extra course credit (M age $= 19.43$). Three were excluded for failing attention checks or responding too quickly on average (mean RT $<$ 1,500 ms), yielding a final sample size of 33 participants.

Design, displays, and procedure. On each trial participants viewed a *target* word at the top middle of the screen and two *option words* positioned symmetrically below. The words were sampled randomly from the pool of 30 with uniform probability on each trial. Participants were instructed to base their judgments either on 1) similarity of associated colors, 2) symbolic similarity, or 3) similarity in "kind of thing" (instructions varied between-subjects). They then selected the option they judged most similar to the target either by pressing the left/right arrow key or by clicking on their choice with a mouse. Words were presented in a font-size adjusted to screen dimensions. There was an intertrial interval of 500ms and no feedback. Each participant completed 200 trials, yielding 2,200 judgments per condition. From these data we used the CrowdKernel ordinal embedding algorithm implemented in SALMON to compute word embeddings in 3 dimensions in each condition.

Denotative Word Embeddings. Two models were used to compute word embeddings for the 30 Experiment 1 concepts. First was a word2vec model trained on the COCA-fiction corpus (79), which was chosen because fiction expresses abstract and analogical language more often than other forms of text found on the internet. The second was GPT-3.5 (48), a transformer-based large language model (LLM) (47). Embeddings were extracted using OpenAI's `text-embedding-ada-002` API, the last model in the GPT series that allowed users to query for token embeddings.

Data, Materials, and Software Availability. Data and code are available at GitHub (80).

ACKNOWLEDGMENTS. This project was supported by NSF Award BCS-1945303 to K.B.S and NSF Award IIS-2007436 to M.G.

Author affiliations: [a]Department of Psychology, Stanford University, Stanford, CA 94305; [b]Department of Mechanical and Industrial Engineering, Northeastern University, Boston, MA 02115; [c]Department of Computer Sciences, University of Wisconsin–Madison, Madison, WI 53706; [d]Department of Psychology, University of Wisconsin–Madison, Madison, WI 53706; [e]Wisconsin Institute for Discovery, University of Wisconsin–Madison, Madison, WI 53715; and [f]Institute of Cognitive Neuroscience, University College London, London WC1N 3AZ, United Kingdom

1. G. Murphy, *The Big Book of Concepts* (MIT press, Cambridge, MA, 2004).
2. S. Laurence, E. Margolis, *Concepts: Core Readings* (MIT Press, Cambridge, MA, 2000).
3. L. W. Barsalou, The human conceptual system. *Camb. Handb. Psycholinguist.* **2012**, 239–258 (2012).
4. G. Lakoff, M. Johnson, The metaphorical structure of the human conceptual system. *Cogn. Sci.* **4**, 195–208 (1980).
5. E. Rosch, C. B. Mervis, Family resemblances: Studies in the internal structure of categories. *Cogn. Psychol.* **7**, 573–605 (1975).
6. E. Rosch, B. B. Lloyd, *Cognition and Categorization* (Taylor & Francis, 1978).
7. M. Kiefer, L. W. Barsalou, "Grounding the human conceptual system in perception, action, and internal states" in *Action Science: Foundations of an Emerging Discipline*, W. Prinz, M. Beisert, A. Herwig, Eds. (MIT Press, Cambridge, MA, 2013), pp. 381–407.
8. T. T. Rogers, J. L. McClelland, *Semantic Cognition: A Parallel Distributed Processing Approach* (MIT press, Cambridge, MA, 2004).
9. T. T. Rogers, "Generalization and abstraction: Human memory as a magic library" in *The Oxford Handbook of Human Memory, Two Volume Pack: Foundations and Applications*, M. J. Kahana, A. D. Wagner, Eds. (Oxford University Press, Oxford, UK, 2024), pp. 172–214.
10. R. L. Goldstone, L. W. Barsalou, Reuniting perception and conception. *Cognition* **65**, 231–262 (1998).
11. M. Kiefer, F. Pulvermüller, Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex* **48**, 805–825 (2012).
12. C. Spence, Crossmodal correspondences: A tutorial review. *Atten. Percept. Psychophys.* **73**, 971–995 (2011).
13. S. Bhatia, Associative judgment and vector space semantics. *Psychol. Rev.* **124**, 1 (2017).
14. T. K. Landauer, P. W. Foltz, D. Laham, An introduction to latent semantic analysis. *Discourse Process.* **25**, 259–284 (1998).
15. S. De Deyne *et al.*, Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. *Behav. Res. Methods* **40**, 1030–1048 (2008).
16. C. E. Osgood, G. J. Suci, P. H. Tannenbaum, *The Measurement of Meaning* (University of Illinois press, Urbana, IL, 1957), vol. No. 47.
17. C. E. Osgood, Exploration in semantic space: A personal diary. *J. Soc. Issues* **27**, 5–64 (1971).
18. C. E. Osgood, Semantic differential technique in the comparative study of cultures. *Am. Anthropol.* **66**, 171–200 (1964).
19. D. R. Heise, *Surveying Cultures: Discovering Shared Conceptions and Sentiments* (John Wiley & Sons, Hoboken, NJ, 2010).
20. C. E. Osgood, W. H. May, M. S. Miron, *Cross-Cultural Universals of Affective Meaning* (University of Illinois Press, Urbana, IL, 1975), vol. 1.
21. J. J. Jenkins, W. A. Russell, G. J. Suci, An atlas of semantic profiles for 360 words. *Am. J. Psychol.* **71**, 688–699 (1958).
22. J. A. Russell, A. Mehrabian, Evidence for a three-factor theory of emotions. *J. Res. Pers.* **11**, 273–294 (1977).
23. G. L. Collier, Affective synesthesia: Extracting emotion space from simple perceptual stimuli. *Motiv. Emot.* **20**, 1–32 (1996).
24. R. D'Andrade, M. Egan, The colors of emotion 1. *Am. Ethnol.* **1**, 49–63 (1974).
25. L. R. Prado-León, K. B. Schloss, S. E. Palmer, *Color, Music, and Emotion in Mexican and US Populations* (New Directions in Colour Studies, Amsterdam: John Benjamins, 2014).
26. S. E. Palmer, K. B. Schloss, Z. Xu, L. R. Prado-León, Music-color associations are mediated by emotion. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 8836–8841 (2013).
27. S. E. Palmer, T. A. Langlois, K. B. Schloss, Music-to-color associations of single-line piano melodies in non-synesthetes. *Multisensory Res.* **29**, 157–193 (2016).
28. K. L. Whiteford, K. B. Schloss, N. E. Helwig, S. E. Palmer, Color, music, and emotion: Bach to the blues. *i-Perception* **9**, 2041669518808535 (2018).
29. F. Samsel, L. Bartram, A. Bares, "Art, affect and color: Creating engaging expressive scientific visualization" in *2018 IEEE VIS Arts Program (VISAP)* (IEEE, 2018), pp. 1–9.
30. X. P. Gao *et al.*, Analysis of cross-cultural color emotion. *Color Res. Appl.* **32**, 223–229 (2007).
31. S. K. Murthy, T. L. Griffiths, R. D. Hawkins, Shades of confusion: Lexical uncertainty modulates ad hoc coordination in an interactive communication task. *Cognition* **225**, 105152 (2022).
32. K. B. Schloss, L. Lessard, C. S. Walmsley, K. Foley, Color inference in visual communication: The meaning of colors in recycling. *Cogn. Res. Princ. Implic.* **3**, 5 (2018).
33. K. B. Schloss, Z. Leggon, L. Lessard, Semantic discriminability for visual communication. *IEEE Trans. Vis. Comput. Graph.* **27**, 1022–1031 (2021).
34. K. B. Schloss, Color semantics in human cognition. *Curr. Dir. Psychol. Sci.* **33**, 58–67 (2024).
35. K. B. Schloss, M. A. Schoenlein, K. Mukherjee, "Color semantics for visual communication" in *Visualization Psychology*, D. A. Szafir et al., Eds. (Springer, 2023), pp. 3–37.
36. K. Mukherjee, B. Yin, B. E. Sherman, L. Lessard, K. B. Schloss, Context matters: A theory of semantic discriminability for perceptual encoding systems. *IEEE Trans. Vis. Comput. Graph.* **28**, 697–706 (2022).
37. M. A. Schoenlein, J. Campos, K. J. Lande, L. Lessard, K. B. Schloss, Unifying effects of direct and relational associations for visual communication. *IEEE Trans. Vis. Comput. Graph.* **29**, 385–395 (2023).
38. K. Mukherjee, A. Mohapatra, T. T. Rogers, K. B. Schloss, Large language models estimate fine-grained human color-concept associations. *Cogn. Sci.* **50**, e70219 (2026).
39. S. Lin, J. Fortuna, C. Kulkarni, M. Stone, J. Heer, Selecting semantically-resonant colors for data visualization. *EuroVis* **32**, 401–410 (2013).

40. A. Jahanian, S. Keshvari, S. Vishwanathan, J. P. Allebach, Colors-messengers of concepts: Visual design mining for learning color semantics. *ACM Trans. Comput. Hum. Interact.* **24**, 1–39 (2017).
41. K. B. Schloss, L. Lessard, C. Racey, A. C. Hurlbert, Modeling color preference using color space metrics. *Vis. Res.* **151**, 99–116 (2018).
42. S. Joshi, S. Boyd, Sensor selection via convex optimization. *IEEE Trans. Signal Process.* **57**, 451–462 (2008).
43. S. E. Palmer, K. B. Schloss, An ecological valence theory of human color preference. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 8877–8882 (2010).
44. J. C. Gower, Generalized procrustes analysis. *Psychometrika* **40**, 33–51 (1975).
45. T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality. *Adv. Neural Inf. Process. Syst.* **2**, 3111–3119 (2013).
46. M. Davies, The corpus of contemporary American English (coca): 560+ million words, 1990-present (2008). https://www.english-corpora.org/coca/. Accessed 12 June 2026.
47. A. Vaswani *et al*., Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30**, 1–11 (2017).
48. T. Brown *et al*., Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **33**, 1877–1901 (2020).
49. K. G. Jamieson, L. Jain, C. Fernandez, N. J. Glattard, R. Nowak, Next: A system for real-world development, evaluation, and application of active learning. *Adv. Neural Inf. Process. Syst.* **28**, e01685 (2015).
50. L. Jain, K. Jamieson, R. Nowak, Finite sample prediction and recovery bounds for ordinal embedding. *Adv. Neural Inf. Process. Syst*. **29**, 2703–2711 (2016).
51. J. J. Singer, K. Seeliger, T. C. Kietzmann, M. N. Hebart, From photos to sketches-how humans and deep neural networks process objects across different levels of visual abstraction. *J. Vis.* **22**, 4 (2022).
52. B. D. Roads, B. C. Love, "Enriching imagenet with human similarity judgments and psychological embeddings" in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2021), pp. 3547–3557.
53. M. N. Hebart, C. Y. Zheng, F. Pereira, C. I. Baker, Revealing the multidimensional mental representations of natural objects underlying human similarity judgements. *Nat. Hum. Behav.* **4**, 1173–1185 (2020).
54. S. Suresh *et al*., "Conceptual structure coheres in human cognition but not in large language models" in *Proceedings of the 2023 conference on empirical methods in natural language processing*, H. Bouamor, J. Pino, K. Bali, Eds. (Association for Computational Linguistics, 2023), pp. 722–738.
55. S. Suresh *et al*., Ai-enhanced semantic feature norms for 786 concepts. *Top. Cogn. Sci.* **18**, e70037 (2026).
56. T. Giallanza, D. Campbell, J. D. Cohen, T. T. Rogers, An integrated model of semantics and control. *Psychol. Rev.* **132**, 1128 (2025).
57. D. Gentner, Structure-mapping: A theoretical framework for analogy. *Cogn. Sci.* **7**, 155–170 (1983).
58. L. W. Barsalou, Perceptual symbol systems. *Behav. Brain Sci.* **22**, 577–660 (1999).
59. C. Kemp, J. B. Tenenbaum, The discovery of structural form. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10687–10692 (2008).
60. R. A. Furrer, K. Schloss, G. Lupyan, P. M. Niedenthal, A. Wood, Red and blue states: Dichotomized maps mislead and reduce perceived voting influence. *Cogn. Res. Princ. Implic.* **8**, 11 (2023).
61. C. A. Thorstenson, A. J. Elliot, A. D. Pazda, D. I. Perrett, D. Xiao, Emotion-color associations in the context of the face. *Emotion* **18**, 1032 (2018).
62. K. B. Schloss, R. M. Poggesi, S. E. Palmer, Effects of university affiliation and "school spirit" on color preferences: Berkeley versus Stanford. *Psychon. Bull. Rev.* **18**, 498–504 (2011).
63. K. B. Schloss, C. C. Gramazio, A. T. Silverman, M. L. Parker, A. S. Wang, Mapping color to meaning in colormap data visualizations. *IEEE Trans. Vis. Comput. Graph.* **25**, 810–819 (2019).
64. S. C. Sibrel, R. Rathore, L. Lessard, K. B. Schloss, The relation between color and spatial structure for interpreting colormap data visualizations. *J. Vis.* **20**, 7–7 (2020).
65. A. Soto, M. A. Schoenlein, K. B. Schloss, More of what? dissociating effects of conceptual and numeric mappings on interpreting colormap data visualizations *Cogn. Res. Princ. Implic.* **8**, 38 (2023).
66. Q. Liu, J. van Paridon, G. Lupyan, Learning about color from language. *Commun. Psychol.* **3**, 60 (2025).
67. J. S. Kim, B. Aheimer, V. Montané Manrara, M. Bedny, Shared understanding of color among sighted and blind adults. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2020192118 (2021).
68. T. T. Rogers, Understanding understanding: Reconnecting theories of human semantic knowledge. *Annu. Rev. Psychol.*, in press.
69. M. A. Schoenlein, K. B. Schloss, Colour-concept association formation for novel concepts. *Vis. Cogn.* **30**, 457–479 (2022).
70. S. L. Franconeri, L. M. Padilla, P. Shah, J. M. Zacks, J. Hullman, The science of visual data communication: What works. *Psychol. Sci. Public Interest* **22**, 110–161 (2021).
71. T. Munzner, *Visualization analysis and design* (CRC Press, Boca Raton, FL, 2014).
72. M. Stone, D. A. Szafir, V. Setlur, "An engineering model for color difference as a function of size" in *Color and Imaging Conference* (Society for Imaging Science and Technology, 2014), vol. 2014, pp. 253–258.
73. D. A. Szafir, Modeling color difference for visualization design. *IEEE Trans. Vis. Comput. Graph.* **24**, 392–401 (2018).
74. C. C. Gramazio, D. H. Laidlaw, K. B. Schloss, Colorgorical: Creating discriminable and preferable color palettes for information visualization. *IEEE Trans. Vis. Comput. Graph.* **23**, 521–530 (2017).
75. J. R. De Leeuw, jspsych: A javascript library for creating behavioral experiments in a web browser. *Behav. Res. Methods* **47**, 1–12 (2015).
76. J. R. de Leeuw, R. A. Gilbert, B. Luchterhandt, jspsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *J. Open Source Softw.* **8**, 5351 (2023).
77. S. E. Palmer, K. B. Schloss, J. Sammartino, Visual aesthetics and human preference. *Annu. Rev. Psychol.* **64**, 77–107 (2013).
78. S. Sievert, R. Nowak, T. T. Rogers, Efficiently learning relative similarity embeddings with crowdsourcing. *J. Open Source Softw.* **8**, e04517 (2023).
79. J. van Paridon, Q. Liu, G. Lupyan, "How do blind people know that blue is cold? Distributional semantics encode color-adjective associations" in *Proceedings of the Annual Meeting of the Cognitive Science Society*, T. Fitch, C. Lamm, H. Leder, K. Teßmar-Raible, Eds. (Cognitive Science Society, Vienna, Austria, 2021), vol. 43.
80. K. Mukherjee, L. Lessard, M. Gleicher, T. T. Rogers, K. B. Schloss, Semantic_spaces_public. GitHub. https://github.com/SchlossVRL/semantic_spaces_public. Deposited 30 August 2026.