Journal of Experimental Psychology: Applied
2001, Vol. 7, No. 2, 112–128

1076-898X/01/$5.00 DOI: 10.1037//1076-898X.7.2.112

Color Coding Information: Assessing Alternative Coding Systems Using Independent Brightness and Hue Dimensions

Kimberly A. Jameson
University of California, San Diego

Jerry L. Kaiwi and Donald Bamber
SPAWAR Systems Center

Can independent dimensions of brightness and hue be used in a combined digital information code? This issue was addressed by developing 2 color-coding systems and testing them on informed and naive participants in signal beam detection and classification experiments for simulated sonar displays. Each coding system's results showed both groups efficiently used encoded information that varied simultaneously along the 2 dimensions of brightness and hue. Findings support the proposed procedures for developing color information codes and the validity of such information codes across different populations. Applied significance of these results is provided by the test of principled methods of color-code construction and the demonstration that extending the information content of user interfaces beyond 1 dimension is feasible in practice.

Information-processing user interfaces, such as air-traffic control systems and automatic banking machines, encode information using salient perceptual attributes such as gradients of brightness or color and varieties of icons and alphanumerics. Much research on color use in digital information displays has focused on error rates and search time for targets (symbols or icons) when color is added as a target feature (Christ, 1975; Donderi, 1994; Nagy & Sanchez, 1990, 1992). Guidelines exist for the number of hues used for targets and backgrounds, the use of luminance and saturation to improve perceptual segregation of colors, and effects due to ambient light conditions (Van Orden & Benoit, 1993). However, there remains some uncertainty regarding the conditions that make color a good choice as an independent, nonredundant coding dimension, and no guidelines exist for the applied use of multidimensional information codes. The present study explored the feasibility of two-dimensional hue–luminance codes and the use of color beyond a simple redundant code in a specific display format in which meaningful information is defined statistically as opposed to iconically. To do this, we pursued three goals: (a) modeling observer performance using ideal-detector theory, (b) developing two-dimensional information codes and meaningful methods of quantizing and encoding information, and (c) conducting empirical tests of implemented codes for gradient intensity signal applications.

Important perceptual considerations arise when constructing a two-dimensional color code that aims to achieve the desired perceptual qualities of independence and equal perceptual increments. Previous work has emphasized issues of stimulus size, surrounding contrast effects, and the selection of color-code values that preserve intended brightness relationships (Carter & Carter, 1988; Kaiwi, Bamber, & Urban, 2000). The most commonly discussed form of perceptual processing in the color-coded user interface literature involves visual search for symbolic representations of discrete categorical data (e.g., Smallman & Boynton, 1990, 1993). A less studied user-interface format involves gradient intensity representations of inherently noisy sensor data (like sonar, radiological scans, or geographical contour map data). This second type of task involves pattern detection and recognition of a signal masked by noise. Processing digitally represented sensor data is perceptually and cognitively quite different from the search tasks involving symbolic representations of categorical data. Data from sensor systems produce digital images that require users to perform visual tasks more akin to pattern recognition than to simple visual search. The present study is concerned with color coding and detection and recognition behaviors for digitally represented sensor pattern data.

When considering pattern recognition for gradient intensity representations, one must also distinguish between (a) situations involving the detection of periodic targets masked by modulated noise (i.e., sinusoidal gratings; e.g., Graham, 1977, 1985; Nachmias & Sansbury, 1974; Thomas, 1985) and (b) those involving the detection of an isolated target of gradient intensity masked by modulated noise. In Situation (a) the detection of near-threshold gratings makes use of the additional information conveyed by the detectable regularity of a target's spatial periodicity. This spatial periodicity, or spatial beat, is considered a *macrofeature* of grating stimuli (Thomas, 1985) and was described by Graham (1985) as a separate perceptual dimension of "spatial phase" or "spatial sym-

Kimberly A. Jameson, Department of Psychology, University of California, San Diego; Jerry L. Kaiwi and Donald Bamber, SPAWAR Systems Center, San Diego, California.

A portion of this work was presented at the 1997 annual meeting of the Association for Research in Vision and Ophthalmology (Beer, Jameson, & Kaiwi, 1997). This research was supported by National Research Council Associateship RV141G, by National Institute of Mental Health Grant NIH R03-MH53126-01, and by a University of California, San Diego, Undergraduate Scholastic Grant No. 43S96 awarded to W. Chappell. We acknowledge the contributions of the members of the Laboratory for Applied Cognitive Research at the University of California, San Diego.

Correspondence concerning this article should be addressed to Kimberly A. Jameson, Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0109.

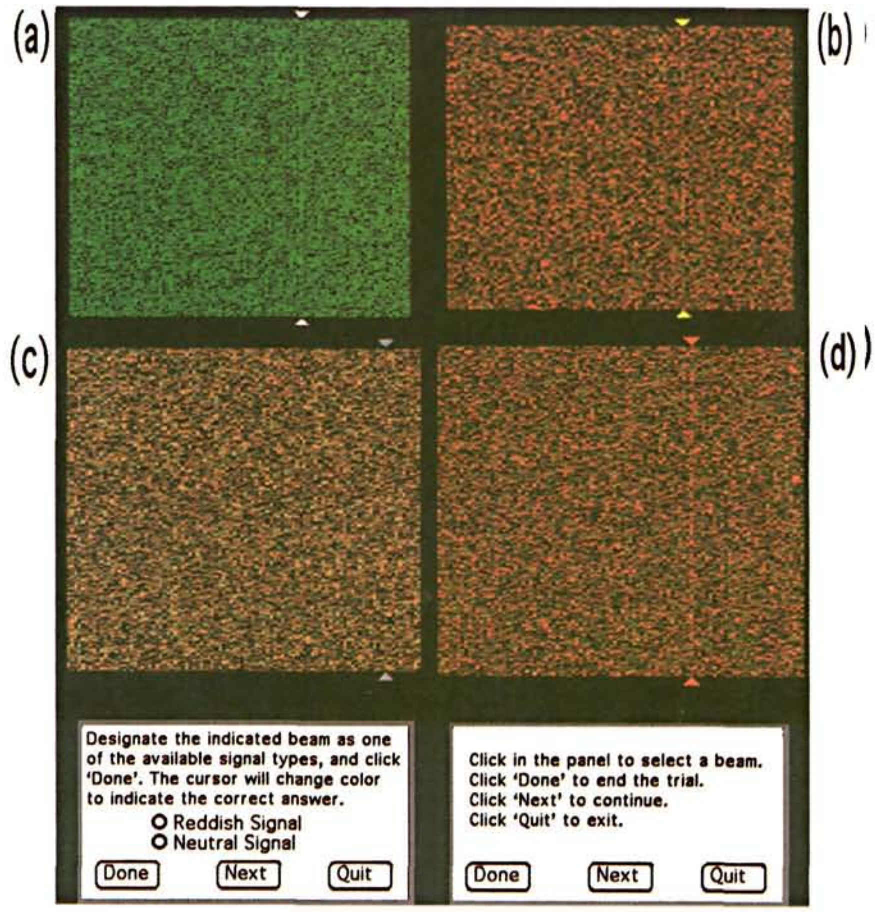

Figure 1. Four examples of stimulus displays representing passive sonar sensor data. Panel (a) represents a monochrome simulated passive broadband data image in which the vertical and horizontal dimensions correspond to time and direction, respectively. Display bins are 1 × 3 pixel arrays. All pixels in a given display bin have the same luminance. There are 250 × 96 display bins—250 sample periods and 96 directions (beams). The image can be updated over time by shifting each row of luminance values down one time interval—thereby dropping out the bottom row—and inserting new luminance values into the top row. Panel (a) also exemplifies an Experiment 1 monochrome test stimulus. All experimental trials consisted of 93 noise beams and a single signal beam, as depicted. (a) illustrates the beam detection task ($d' = .41$) with a candidate signal beam already selected by the participant and highlighted with top and bottom markers. In Experiment 1 this stimulus panel was shown in conjunction with the white response-format box shown in Panel (d). Once the participant was satisfied with a selection, the handheld mouse was used to click the "done" button, and feedback was shown concerning the location of the actual signal beam by the presentation of a cursor indicating the true signal beam. Presentation of the next trial was initiated by a mouse click and was self-paced. Panel (b) represents Experiment 2's brightness code with chromatic noise (encoded here by Code A). The task-and-response format is identical to that described for Panel (a), which uses the response-option format illustrated in Panel (d). Panel (c) shows Experiment 3's hue code with brightness noise (encoded here by Code B). As in all experiments, each display consisted of 94 columns and 250 rows of 1 × 3 pixel bins. Each bin was assigned one hue-luminance value. The experiment involved 1,504 trials per participant. In the Experiment 3 stimuli the chromaticity index and the luminance index were uniformly distributed across noise display bins. Signal display bins also had uniformly distributed luminance indices, but compared with noise display bins, signal chromaticities were on average in the hue gradient direction of red. As illustrated in the white response-option window, the task differed from that presented in (a), (b), and (d) in that observers were asked to classify a predesignated beam into a "reddish" or "neutral" signal category. (d) represents Experiment 4's combined brightness and hue code (encoded here by Code A). The task performed was identical to that in Experiments 1 and 2—(a) and (b), respectively.

metry" (p. 1469) distinct from the spatial location of a signal in two orthogonal dimensions. Empirical studies show that observers use spatial beat patterns when identifying stimuli (e.g., Thomas, 1985). By comparison, in Situation (b) signal information is represented simply as an isolated target in noise. Situation (b) is the subject of the present study and involves different detection behavior from Situation (a), which has been studied elsewhere.

Situation (b) is exemplified by U.S. Navy visual images used to analyze passive sonar data. Passive sonar "listens" to underwater sound intensity and uses digital signal processing to construct target corridors, or "beams" of sound intensity samples, for a specified underwater area. Passive sonar data parameters are time, direction, and intensity. Visual representation of beam data maps time and direction onto the two-dimensional CRT display and quantizes sound intensity with display pixel luminance.

Figure 1a depicts a CRT representation of passive broadband azimuthal data. Time-sampled data are quantized into luminance levels (at constant chromaticity), and each green-phosphor luminance level is calibrated to a constant increase in perceived brightness. Pixel luminance is thus mapped as a monotonic increasing function of sound intensity. Luminance values in a signal beam are skewed toward higher values, and the rest of the image has a uniform distribution due to the quantization procedure. Modifying displays such as that presented in Figure 1a so that more information can be presented would be an important advance. In such a two-dimensional system a single coding value could convey two qualitatively different kinds of information—say, a hue-encoded semantic value (e.g., a threat index) and a brightness-encoded signal power, or certainty, index—thus allowing more information to be packed into displayed elements. However, if the color coding was not done carefully, adding new information would very likely complicate the ability to extract information about acoustic energy from the display. In displays such as Figure 1a it is easy to compare two vertical beams and determine which one represents more acoustic energy when only brightness encodes more signal (and hue and saturation are kept uniform). Devising a luminance code that also uses different hues and saturations to encode information is a more difficult task. Here we evaluate the feasibility of such a use of luminance and chromaticity to simultaneously encode two independent sources of data. The experiments address two basic questions: First, can color be combined with a brightness code without detrimentally affecting the encoded signal, and, second, can color be used as a separate coding dimension when conjoined with brightness-encoded information? Positive answers to these questions would suggest that a multidimensional coding system is a practical coding alternative.

Applied situations dealing with the visual representation of gradient sensor data typically encounter three challenges: (a) the determination of how to model the observer's visual detection performance, (b) the practical and perceptually optimal assignment of discrete coding levels to continuous gradient-sensor data, and (c) the development of principled procedures for constructing coding systems that are effectively processed by observers and that can be extended and generalized across display formats. Below we address these issues for representations of acoustic energy data from underwater passive sonar sensors.

Information Code Development and Stimulus Construction

The stimuli used in the present experiments include two related levels of specification. One level involves the human observer theory underlying the quantization of sensor data, and the second level involves visual perception considerations regarding specific *color-code* appearances used to represent the quantized levels of information. We used Green and Swets' (1967) Theory of Signal Detection (TSD) to quantify visual detection performance and to define a luminance code. One advantage of the TSD approach is the mathematical convenience of the ideal-detector model for Gaussian noise, which was used in the present research in the construction of the stimuli and the experimental designs (see Green & Swets, 1967, Part II). Even though the intensity of broadband sound is demonstrably not Gaussian, it is arguably justifiable and useful from a perceptual point of view to "whiten" the signal by averaging a large number of intensity sample values before the data are encoded for display. If the number of samples is large, then average sonar beam intensity values will have an approximately normal distribution. Consequently, we used a Standard Normal distribution to simulate the statistics of averaged sonar beam outputs. Provided reasonable simplifying assumptions, the relationship between the Green and Swets detection index and white noise is given by $d' = (s_o/n_o)\sqrt{w \times t}$, where s_o is signal density, n_o is noise density, w is bandwidth, and t is sample interval (see Green & Swets, 1967, pp. 174–175). With the Standard Normal Model, we were able to conveniently construct simulated sonar images for both forced-choice testing and yes–no testing.

In the present experiments we assumed that the value of d' strictly determines optimal performance (i.e., the performance of an *ideal detector*). Because human observer performance can be mathematically related to an equivalent d', it is common usage to refer to d' as the *sensitivity* of an observer. Therefore, in order to avoid confusion, we use d' to denote the sensitivity of the mathematically optimal observer and d'_o to denote the sensitivity derived from observed data. We refer the reader to Green and Swets (1967) for a formal explanation of how various types of detection performance can be defined in terms of d'.

Quantization of Gradient Intensity Data Using TSD

Figure 2 illustrates the classical statistical assumptions of the TSD used in these experiments and the relationship between d' and a three-bit visual quantification scheme for constructing monochromatic images such as Figure 1a. The noise distribution and the signal distribution differ only by their mean values (equal to zero and d', respectively) with both distributions' standard deviation equal to one.

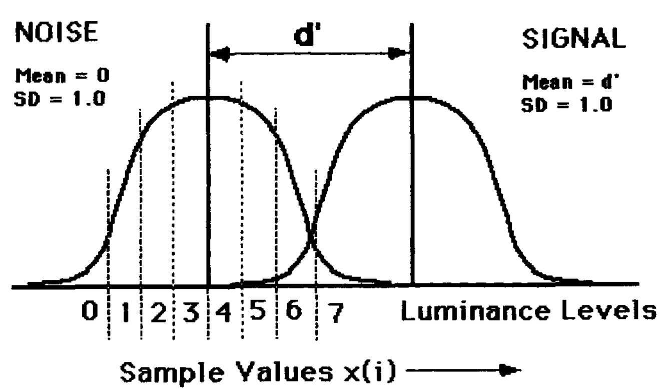

Figure 2. Standard Normal distributions are used to quantify broadband sonar data for digital displays. The Noise distribution on the left is N(0, 1), and the Signal distribution is N(d', 1). The vertical dashed lines indicate cutoff criteria for assigning luminance-level values to display bins.

The vertical dashed lines in Figure 2 indicate cutoff criteria for assigning luminance-level values to display bins. (Display bins are the individually controlled information units—1 CRT pixel high × 3 pixels wide in size—that make up a vertical signal beam.) The selection of the cutoff between a luminance level of zero and of one determines the probability (or "marking density") that any particular display bin will be assigned a luminance value greater than the minimum value. Luminance values for display bins are based on sample values from either the noise distribution or the signal distribution. Suppose cutoff values are selected so that the region under the noise curve is partitioned into equal areas. In this case samples from the noise distribution will be quantized with luminance values that have a uniform distribution. In contrast, samples from the signal distribution will tend to be quantized with luminance values that clump at the high end. Figure 1a exemplifies an image containing one signal displayed at a marking density of 87.5%. In the signal beam, indicated by arrow pointers, higher luminance values occur with probabilities that increase with the value of d'. Using this quantization method, we constructed a one-dimensional luminance code for use in Experiment 1. Experiments 2, 3, and 4 below use two-dimensional color codes.

Monochrome Stimuli

Monochrome images (like Figure 1a) were constructed using the model depicted in Figure 2. Cutoff values were selected that would divide the area under the Normal Distribution into eight equal parts corresponding to eight luminance levels. Using the green CRT phosphor, the monochrome code values ordered from dark to bright were 0.2, 2.8, 6.5, 11, 16.4, 22.2, 28.7, and 35.6 cd/m^2. In each image, there is one vertical signal beam, and all other columns represent noise. All display-bin luminance values in the signal beam were derived from the signal distribution, whereas the noise display bins were determined by generating a random number from an N(0, 1) distribution and selecting the index number between 0 and 7 (inclusive) corresponding to the cutoff interval containing the random number. Figure 3a schematically depicts the relation of the sampled data distributions to the coding dimensions. The index number was then used to select a luminance value: The higher the index number, the higher the corresponding luminance. The luminance value of a signal display bin was determined by first generating a random number from an N(0, 1) distribution. The signal d' value was then added to the random number and the sum quantized by selecting the index number between 0 and 7 (inclusive) corresponding to the cutoff interval containing the sum. The index number was then used to select a luminance level. In the monochrome code all display bins equaled the chromaticity of the CRT green gun.

Constructing Two-Dimensional Color Codes

We aimed to construct two-dimensional color-coding systems that would (a) yield detection performance on a par with that observed for a one-dimensional brightness code and (b) provide generalizable procedures for developing two-dimensional color codes. As with the monochrome stimuli, two-dimensional color-code stimuli also used vertical signal beams composed of individual CRT display bins.[1] Because display bins subtend < 2° visual angle, it was assumed that the brightness code used in a display bin is a function of luminance only. Our application required that eight levels of brightness be implemented to represent sonar signal strength; we also defined a perpendicular dimension with eight hue levels to encode eight levels of semantic information. This two-dimensional code could use the strength dimension to represent levels of certainty of information, whereas the semantic information dimension might identify categorical information, such as hazard for a red code and safety for a green code. In addition, observers should be able to compare two display bins and, on the basis of their relative brightnesses, determine which display bin represented the larger value for the acoustic energy parameter even if the two display bins had different hues and saturations. Two such color codes are described below.

Code A

One commonly practiced method of color coding is to assign digital display increments across color categories to represent measured increments in the available data (e.g., some medical image scans). Figure 4 presents a color code based on this idea of using physical attributes of CRT digital increments (hereafter abbreviated Code A).[2] Code A has eight levels (rows) of luminance, each composed of eight hues ranging from green through yellow to red (columns). All hues had the maximum saturation that the CRT could deliver. Values for approximately equal steps in the fundamental luminance and hue gradients were based on digital phosphor values. Each of the remaining elements of Code A was designed to have (a) rows with elements of the same luminance and (b) columns with elements of the same chromaticity. Code A's 8 × 8 color values assume that, to a reasonable approximation, brightness is proportional to the cube root of luminance (Wyszecki & Stiles, 1982). For display design this assumption has been found justifiable in practice (Widdel & Post, 1992). Code A is discussed further in Kaiwi et al. (2000).

Code A's construction method is a commonly used and efficient procedure by which to derive a color-coding scheme in that it uses a simple monotonic model for the perceived relationships between physical luminance and subjective brightness and between brightness and hue. However, some problems interpreting such a code might arise when two display bins have different hues and saturations and the subjective perception of relative brightness may be highly ambiguous. Even though one display bin has a substantially greater luminance than the other, an observer may feel unsure which display bin is brighter. To address these issues, we compared performance under Code A with that under an alternative code (hereinafter Code B) developed with subjective brightness nonlinearities in mind.

Code B

A second *psychologically based* color code, Code B, explicitly addresses potential problems with variation in brightness and hue perception that may arise under Code A. Code B was constructed to minimize known perceptual nonlinearities that might impact the

[1] Recall that display bins are 1 CRT pixel high by 3 pixels wide.

[2] Note that all figures we present to depict color codes and stimuli only approximately reproduce the CRT colors used in our actual experiments.

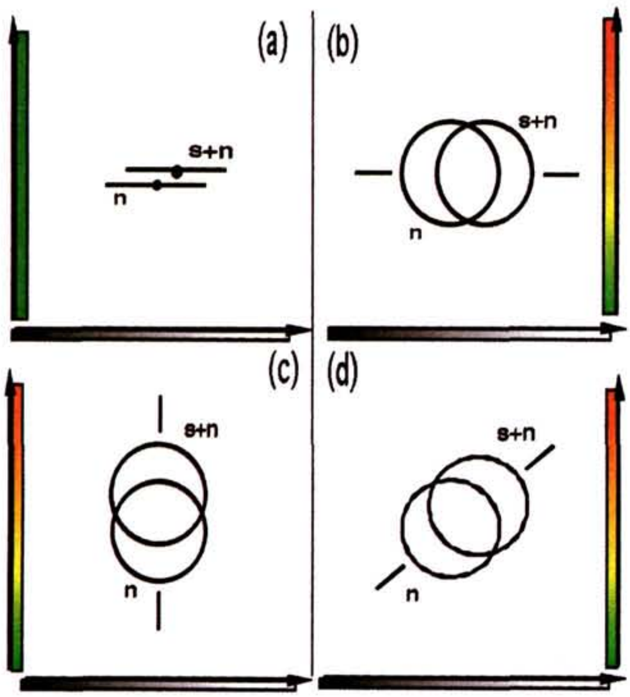

Figure 3. Schematic representations of the encoding variations used in Experiments 1, 2, 3, and 4. Each panel presents the statistical relationships underlying the signal and noise distributions from which coding indices were randomly selected for the construction of experimental stimuli. The independence and equal variance depicted in the panels model the statistics of the underlying data rather than the distribution of perceptual units inherent in the stimuli used. (a) depicts the theoretical relations between signal (s) and noise (n) in the one-dimensional monochrome brightness code in Experiment 1. (b) depicts the same for the two-dimensional hue and brightness relations used in Experiment 2's luminance-encoded signal condition. (c) depicts the two-dimensional hue and brightness relations used in Experiment 3's chromaticity-encoded signal condition. (d) depicts the two-dimensional hue and brightness relations used in Experiment 4's integrated brightness- and chromaticity-encoded signal condition.

effectiveness of a coding scheme using combined hue and brightness to represent information.

To construct a Code B that simultaneously accounts for perceptual and psychological characteristics inherent in human operators, we used the *Munsell Book of Color,* which is a color-ordered system that is based on empirically assessed human color perception (Newhall, Nickerson, & Judd, 1943). Code B aims to account for quirks of the human perceptual system that cause nonlinear relationships between the physical increments of light and the subjective color appearances (e.g., Bezhold-Brücke and Abney shifts). The Code B method also represents a generative approach to constructing new coding systems or extending the coding levels of existing systems. Being based on perceptual data, the method yields nonarbitrary coding systems that should optimize operators' processing of a code.

The Munsell color space uses standard perceptual dimensions (i.e., hue, lightness, and saturation, referred to as Hue, Value, and Chroma, respectively). The method for color code construction is depicted in Figure 5a, where a curved plane of perceptually constant saturation and radial constant-hue loci are embedded in a stylized Munsell solid. Code B's 64 color-coding values are determined by intersections of the colored lines with the curved plane for eight indicated lightness levels (Value = 2 to 9). The contour indicated by hash marks at Chroma level = 6 intersects the eight radial hue curves at points of perceptually constant saturation and brightness. Loci of constant hue and constant saturation (i.e., Chroma) at a fixed level of brightness (i.e., Value) are reprinted in Wyszecki and Stiles (1982, pp. 840–861). This heuristic yields the 64 coding values approximated in Figure 5b. This procedure automates and improves color-code construction and has the advantage of being based on psychological constants in a perceptual color space. Code B also provides information regarding the effects of strictly uniform saturation. That is, whereas Code A used heterochromatic saturations produced by maximized display phosphors, Code B explicitly constrained saturation levels for all coding values to a level known to maximize linearity of the brightness–saturation relationship. Controlling saturation in this way yielded Code B's 64 coding levels with an average luminance twice that of Code A. See Kaiwi et al. (2000) for further details on Code A and Code B properties and construction methods.

Stimuli Color Coded With Code A and Code B

Color-coded stimulus images (see Figure 1, b–d) were constructed by computing an index number for chromaticity (as done for luminance) in addition to an index number for luminance (described above). The resulting number was used to select a row in either Code A (see Figure 4) or Code B (see Figure 5b). Thus, both luminance and chromaticity indices have uniform distribu-

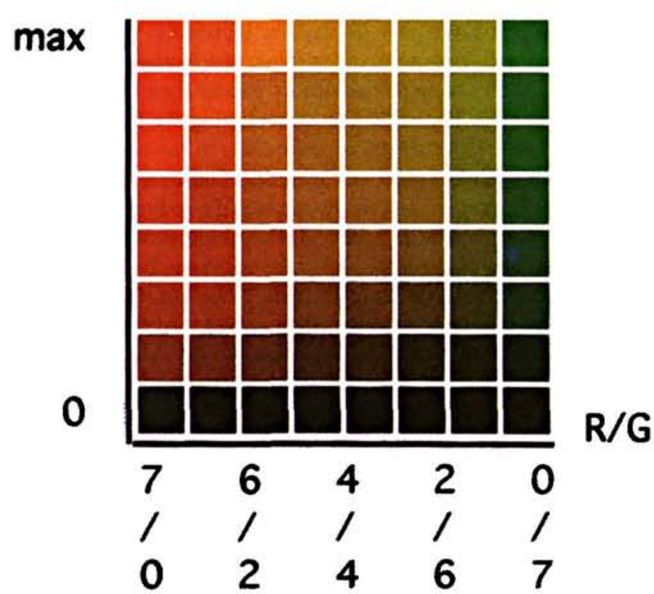

Figure 4. Schematic representation of the Code A color code. For the brightest red (R) value, the red gun was maximized. The brightest green (G) in the table was then set to match. Remaining hues were calculated by using a ratio of red and green gun activation. Brightness steps were made approximately proportional to the cube root of luminance. The 64 values of Code A were measured as Commission Internationale de l'Eclairage (1931) average values $x = 0.425$ ($SD = 0.116$), $y = 0.426$ ($SD = 0.072$), and $Y = 4.828$ ($SD = 4.012$).

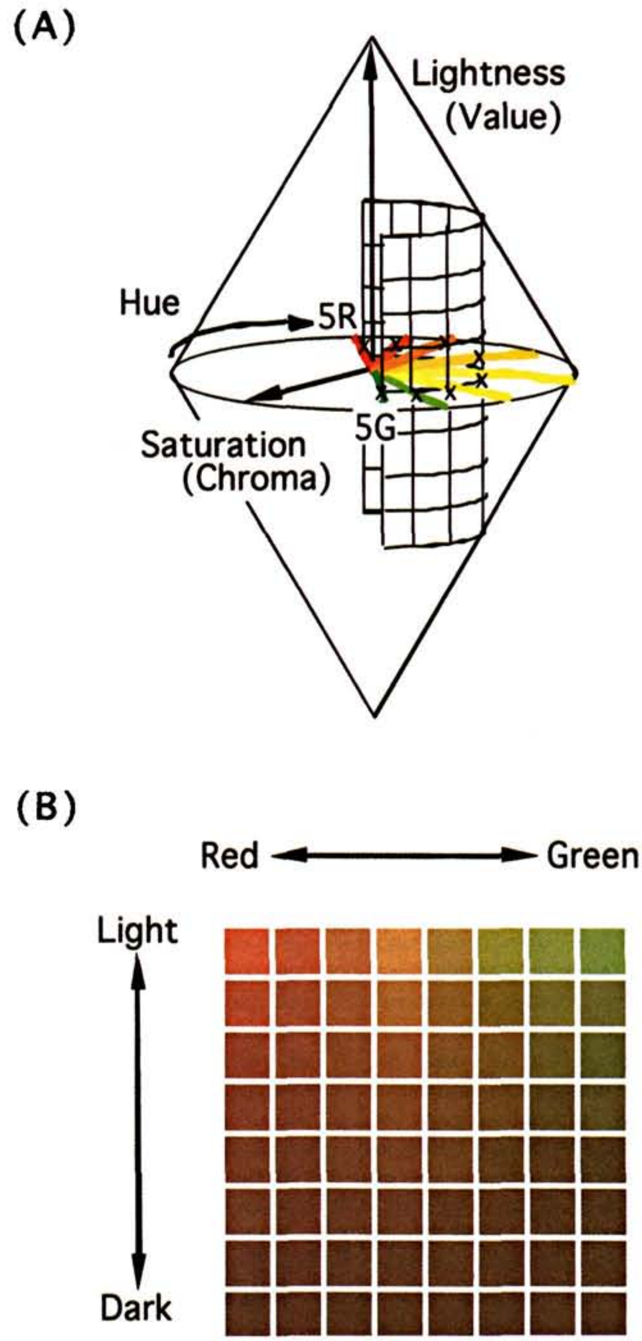

Figure 5. Schematic representation of the Code B color code. (a) The heuristic used eight equally stepped Munsell hue loci from 5G (green) to 5R (red) represented here as colored lines. For each hue, eight equally stepped brightness levels (Munsell Value 2 to 9) were selected (saturation held constant at Munsell Chroma = 6.). An approximation of the Code B 8 × 8 function appears in (b). The 64 values of Code B were measured as Commission Internationale de l'Eclairage (1931) average values $x = 0.384$ ($SD = 0.062$), $y = 0.399$ ($SD = 0.046$), and $Y = 9.480$ ($SD = 7.430$). Because Code B is based on constant-hue loci from the Munsell color space, fitting the uniform-saturation color code within the CRT gamut necessitated the selection of a relatively midrange saturation portion of the Munsell color space, hence the more desaturated appearance of Code B.

tions across noise display bins, whereas signal display bins tend to be brighter and/or redder depending on the test condition as illustrated in Table 1. How will such two-dimensional stimuli be processed by observers?

According to Maddox (1992), "a fundamental issue in human perception is to determine how different stimulus dimensions interact during perceptual processing" (p. 147). Garner (1974) explained that pairs of stimulus dimensions are separable if people

Table 1
A Summary of the Color-Code Index Distributions in Experiments 1–4

	Luminance		Chromaticity		
Experiment and sample	Noise bins	Signal bins	Noise bins	Signal bins	Task
Experiment 1, DOD only	Uniformly distributed	Biased toward bright	Constant green	Constant green	Forced choice
Experiment 2, DOD & UCSD	Uniformly distributed	Biased toward bright	Uniformly distributed	Uniformly distributed	Forced choice
Experiment 3, DOD & UCSD	Uniformly distributed	Uniformly distributed	Uniformly distributed	Biased toward red	Yes–no
Experiment 4, DOD & UCSD	Uniformly distributed	Biased toward bright	Uniformly distributed	Biased toward red	Forced choice

Note. DOD = Department of Defense; UCSD = University of California, San Diego.

can selectively attend to one or the other at will without interference from the unattended property. The implication is that the perceptual representations are independent. Pairs of dimensions are integral if people cannot selectively attend to one without also perceiving the other. For example, saturation and lightness are two dimensions that seem to be processed together. As discussed by Kadlec and Townsend (1992), Garner's original concepts of *separability* and *integrality* have raised some confusion with regard to their exact meaning. For the present study we use a definition following Tversky and Krantz (1970), which simply defines *perceptual dimensions* as the organizing principles or factors along which stimuli are perceived and structured. Here we examine perceptual dimensions that correspond in a straightforward manner with psychophysical dimensions of luminance variation (encoding signal intensity) and chromatic variation (encoding signal type). The present study differs from studies of perceptual dimensionality in one very important respect: We did not aim to investigate whether the two dimensions used are perceptually *integral* or *separable,* rather, we simply aimed to investigate pragmatically how the two stimulus components influence observer performance when they are evaluated as isolated information-coding dimensions compared with when they are judged together as a composite coding dimension.

What empirical performance issues bear on such two-dimensional color-coded stimuli? We assume that information-processing losses associated with a two-dimensional color code may occur in two basic ways. Following Hershman, Kaiwi, and Pilmore (1987), performance under luminance-code processing (Experiments 1 and 2) is expected to be inferior to that predicted for an ideal observer. Small-field color differences were expected to produce information losses under a chromaticity code (Experiment 3), and color-code performance losses were predicted to be greater than those associated with a monochromatic luminance code (cf. Widdel and Post, 1992, pp. 19–20). Also, consistent with Garner's (1974) concept of *orthogonality loss* (p. 126), noise in one channel may mask information in another channel (Experiments 2, 3, and 4). Thus, compared with monochromatic encoding (Experiment 1), luminance encoding masked by chromatic "noise" may yield relatively poorer performance (Experiment 2), and chromatic encoding masked by luminance noise may also yield relatively poorer performance (Experiment 3). Furthermore, destructive interference during the perceptual integration of simultaneously processed luminance- and chromaticity-encoded signals (Experiment 4) may result in losses that are independent of losses possibly incurred when one channel produces only noise (Experiment 2 and 3). Our aim in the present experiments was to isolate and measure these sources of information loss and to test the notion that increasing information-processing demands by the introduction of composite luminance–chromaticity coding results in decreasing detection performance.

Experimental Studies

General Method

Experiments assessed two participant populations, using two different computer display systems. These participants were from a Department of Defense (DOD) Navy laboratory in San Diego, California, and from the human participant pool at University of California, San Diego (UCSD). Five of the DOD participants took part in Experiment 1, whereas Experiments 2, 3, and 4 included 4 of the participants from Experiment 1 plus an alternate 5th DOD participant (Jameson). The UCSD sample consisted of 17 naive college undergraduates from UCSD and an informed author (Jameson) also tested in the DOD experiments. The college undergraduates participated voluntarily and received partial course-credit compensation, plus a nominal monetary reward for correct detections. Experiments 2, 3, and 4 all assessed the same 17 undergraduate participants. The UCSD undergraduate participants were completely naive to the purpose of the experimental studies and participated in the three experiments in one of six randomly assigned sequential orders. The order for DOD participants was not strictly regulated. All participants had normal or corrected visual acuity and normal color vision (per the Farnsworth–Munsell 100 Hue Test and/or the Ishihara Pseudo-Isochromatic Plates, Ishihara, 1994).

All DOD participants were assessed using computer-controlled stimuli presented on a 19-in. (48.3-cm) RasterOps CRT display (Model 1960 with a Trinitron GDM 1950 tube) corrected for phosphor nonadditivity (Brainard, 1989). All UCSD participants used a 15-in. (38.1-cm) Magnavox CRT display (Trinitron Model CM2080GY01) similarly corrected for phosphor nonadditivity. Viewing position and the physical environment were controlled across all experiments: 18–18.5-in. (45.7–47-cm) viewing distance and ~2.0 Lux ambient illuminant. Calibration of both devices was maintained throughout all experiments reported. Prior to participating in the experiment, participants were dark adapted to the ambient illumination for approximately 10 min. Observers completed a large number of practice trials to ensure that performance had reached asymptote. Prior to each session a minimum of 20 practice trials were completed. All trials were unspeeded and self-paced. Each experiment consisted of a different random sequence of 376 judged images.

In all experiments each stimulus image was formatted as in Figure 1 in a 96-column by 250-row array of display bins. This stimulus subtended an approximate visual angle of 11.5° vertical by 13.7° horizontal. The extreme left and right columns of the image (i.e., columns 1 and 96) were never used as signal beam locations to avoid crispening effects (Wyszecki & Stiles, 1982, p. 497). Thus, 94 different stimuli consisting of exactly one signal beam and 93 noise columns were assessed. Because no agreed-on model exists relating detection performance directly to color perception, we used TSD techniques to quantify and compare human observer forced-choice and yes–no performance to the performance of a mathematically defined optimal observer derived from the statistics of the input data. The luminance information content of the signal beam in these 94 panels varied from the surrounding luminance noise dependent on a specific d' level of the signal. A signal display bin visually encoded data sampled from a signal distribution, and a noise display bin visually encoded noise distribution data. All display bins were either monochrome or color coded depending on the test condition. These display properties were identical for all experiments described.

Experiment 1 (monochrome encoding) used signal and noise data encoded solely by luminance, thereby providing baseline monochromatic signal-detection performance. Ninety-four possible beam locations and four possible d' signal levels were tested, yielding 376 different stimulus panels. The four d' levels assessed were .166, .241, .291, and .410 (Hershman et al., 1987). These d' levels represent the information content contained in a display bin assuming an ideal signal-detection model. That is, information content increased with increases in d'; thus, beam detection performance was predicted to monotonically increase with increasing d' levels. (Signal beam d' is defined later.) Only DOD participants were assessed under the monochrome luminance code as measured by the observed d' for detection performance (dependent variable). Luminance indices in signal display bins varied according to a distribution biased toward high indices, whereas luminance indices in noise display bins had a uniform distribution. Chromaticity was constant and equaled the CRT green phosphor.

Experiment 1 used monochrome (green) data images and four d' values, and Experiments 2, 3, and 4 used data images encoded with Code A or Code B (50% of the time each) and only two display-bin d' values (.241 and .291).[3] Thus, Experiments 2, 3, and 4 used a 2 × 2 design for the variables color code (Code A and Code B) and participant variation (naive–UCSD vs. informed–DOD participants), and detection performance was assessed by comparing observed d' values (as dependent variables).

Experiment 2 (luminance signals masked by chromatic noise) evaluated the masking effects of chromatic noise on the detection of luminance-encoded signal data. As in Experiment 1, signal information was conveyed entirely by luminance; however, chromaticity indices varied randomly across display bins according to a uniform distribution to simulate chromatic noise. In Experiment 2 participants located the signal beam presented in each stimulus panel. Instructions were to search for the "brightest vertical beam." All other aspects of Experiment 2 were identical to those of Experiment 1.

In Experiment 3 (chromaticity-encoded signals masked by luminance noise) signal information was conveyed entirely by chromaticity (i.e., chromaticity indices in signal display bins had a distribution biased toward red, whereas chromaticity indices in noise display bins were uniformly distributed). Display-bin luminance indices varied according to a uniform distribution. The task in Experiment 3 required participants to classify indicated signal beams into one of two categories (discussed further below). The 376 stimuli used in Experiment 3 were identical to those presented in Experiment 2 with the following exceptions: (a) Entire stimulus panels (i.e., signal beam and surrounding beams) incorporated irrelevant luminance noise and meaningful chromatic content, and (b) the experimental task was a yes–no beam classification task rather than the 94-alternative forced-choice signal-detection task used in Experiment 2.

In Experiment 4 (luminance–chromaticity integration), signal information was conveyed by both luminance and chromaticity. The task was identical to the beam detection task of Experiments 1 and 2. As in Experiments 2 and 3, two color-coding schemes were tested in 94 different beam locations at two possible d' signal levels, yielding 376 display panels that were judged by each participant in the experiment.

A summary of index distributions and the corresponding task in each experiment is given in Table 1. Data from these experiments are analyzed to compare (a) average performance between experiments, (b) average performance across the two participant populations sampled, and (c) average performance under the two color codes assessed. All reported tests of significance used two-tailed paired Student's t tests unless otherwise stated.

Experiment 1: Monochrome Encoding

Experiment 1 established baseline performance measures for the detection of signals coded in monochrome data images. Figure 3a schematically depicts the monochrome code as it varies in brightness relative to an ambient noise signal. This is also shown in Figure 1a, where the brightness level of the indicated signal beam is identifiable within noise. Experiment 1 results provide a performance baseline to compare with results from color-coded displays to determine whether the introduction of irrelevant color impairs baseline detection performance.

Method

Participants. Participants were 5 paid DOD participants as described above.

Materials. Figure 1a shows a typical Experiment 1 test image (d' = .41) with response format shown in Figure 1d. A computer algorithm generated 94 stimuli for each of four d' values, yielding a total of 376 test images. Participants judged 94 of these vertical beam images for each d' at each possible signal beam location. Physical characteristics of the stimuli and device were as described above.

Design. Experiment 1 trials required participants to perform a 94-alternative forced-choice task. Each participant was presented with a dif-

[3] The two d' levels used in Experiments 2, 3, and 4 were .241 and .291. d' levels .166 and .410 were previously assessed in color-code pilot experiments and were found to be close to the floor and ceiling of detection, respectively. For this reason, and following the reasoning that the number of trials in the monochrome experiments should equal that in the color-code experiments to minimize possible learning or fatigue effects, the extreme d' values were omitted from the color-code experiments. This decision was also influenced by real-world beam detection considerations.

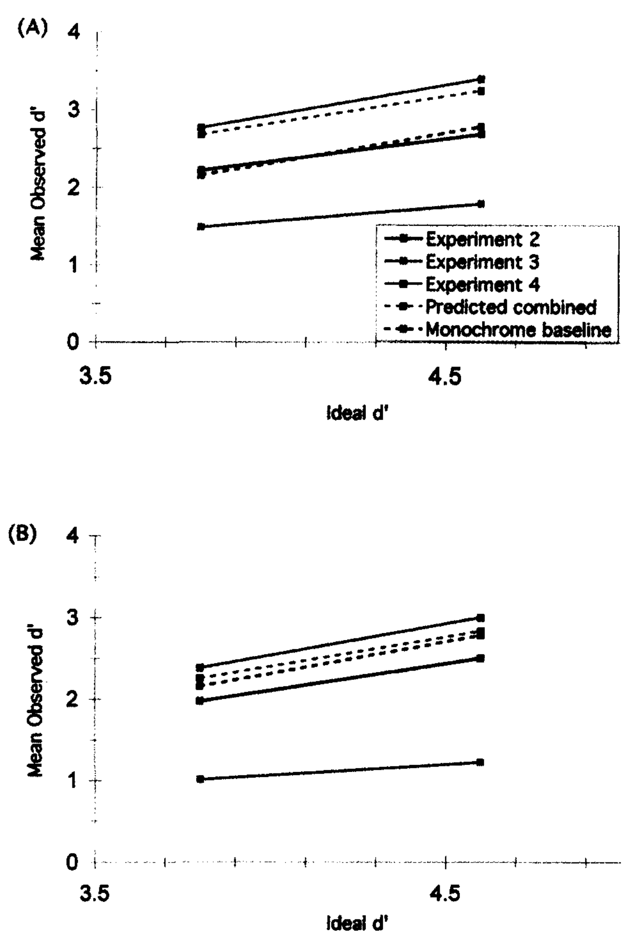

Figure 6. Averaged task curves for (a) Department of Defense (DOD) participants and (b) University of California, San Diego, participants (with the DOD's Experiment 1 monochromatic baseline performance indicated by the black dotted line). For these analyses the mean observed d' values represent an averaging of results across color-coding schemes. Comparisons of the average d'_o values across the different tasks used in Experiments 2 and 4 against the average d'_o values found for Experiment 3 require an assumption about the formal model underlying the tasks. Because of the underlying signal-detection model used in the design of the present experiments, it is meaningful to compare d' results across the forced-choice and yes–no tasks used. That is, because acoustic signals are typically Gaussian distributed we can use a strong version of the theory (making the assumption of a Gaussian-distributed signal with known variance), and, as a result, we know that the ideal observer in the case of the forced-choice and yes–no test are formally related (especially because large samples of the signal were used in the stimulus displays; see Hershman, Kaiwi, & Pilmore, 1987, for a discussion). These data support the notion that brightness codes seem to be more salient perceptually in that naive and informed observers alike perform closer to a theoretical ideal when interpreting brightness codes, compared with chromatic classification code performance. Thus, the main finding arising from a comparison of the results from Experiment 2's 94-alternative forced-choice task and Experiment 3's yes–no task is that chromatic signals are more difficult to process than luminance signals, although both kinds of signals were found to effectively aggregate across display elements of very small field size. This latter result is an important demonstration that gradient information can be effectively aggregated and processed when the display representation uses information units of a spatial extent well below 2° visual angle.

ferent randomized sequence of 376 stimulus images. The number of stimuli judged during the experimental session was controlled by the participant.

Procedure. Participants were asked to examine each stimulus image, compare each beam from top to bottom and find the "brightest" vertical beam, and "mark" that signal beam by moving a mouse-controlled cursor and clicking the mouse. In general, Experiment 1 instructions for signal beam locating strategies were not elaborate because all the DOD participants were acquainted with the task. Having selected a candidate "signal" beam, participants were allowed to modify their selection as often as desired.[4]

Results

Although Experiment 1 assessed four d' levels, we present only data from two d' levels as baseline measures for d' levels used in Experiments 2, 3, and 4. Performance in Experiment 1's detection task is based on the participant's ability to visually "integrate" across an entire vertical beam, composed of 250 stacked display bins. It is easy to show that given n independent display-bin d' values, the optimal mathematically integrated d' for the entire beam is as follows:

$$d'_{\mathrm{I}} = \sqrt{\sum_{k=1}^{n} d'^2_k}, \tag{1}$$

where k is the index of display bins in a column. Green and Swets (1967) discussed both the integration model (pp. 238–239) and the mathematical relationship between d' and proportion correct in a forced-choice test (pp. 45–50).

If the value of d' happens to be constant for all display bins, then the above equation simplifies to the following:

$$d'_{\mathrm{I}} = \sqrt{n} \times d'. \tag{2}$$

Thus, given 250 display bins per beam, the integrated beam d'_{I} value equals the product of 15.8 and the pixel d' value. Table 2, row 1, gives the beam d'_{I} values corresponding to display-bin d' values (.241 and .291), as well as Experiment 1's empirically observed d'_o values inferred from the average observed proportion correct. Experiment 1 showed that average d'_o values from 5 DOD participants equaled $\mu(d'_{o.241}) = 2.165$ and $\mu(d'_{o.291}) = 2.835$. This performance is comparable to that in Hershman et al. (1987), suggesting the Experiment 1 performance is typical of that seen for monochrome-encoded data in a static beam detection task. Experiment 1's d'_o values serve as baseline measures in the analyses below.

Experiment 2: Luminance Signal Masked by Chromatic Noise

Given the baseline monochrome detection performance measured in Experiment 1, we next sought to quantify detection performance for a luminance-encoded signal masked by randomly varying chromaticity. The perceptual issue is the degree to which luminance processing is independent of chromaticity processing

[4] In all experimental tasks presented here, pilot studies showed that feedback of this sort produced no appreciable learning or local practice effects after completion of the required practice trials.

Table 2
Average Observed d′ Values for All Experimental Conditions Assessed

Experiment and condition	Encoding	Signal beam d' levels tested			
		DOD sample		UCSD sample	
		3.81	4.60	3.81	4.60
Experiment 1 baseline	Monochrome	2.16	2.83	—	—
Experiment 2, luminance encoding	Code A	2.20	2.69	1.93	2.47
	Code B	2.24	2.69	2.02	2.55
Experiment 3, chromaticity encoding	Code A	1.61	2.06	1.32	1.53
	Code B	1.36	1.53	0.71	0.92
Experiment 4, TSD predictions	Code A	2.73	3.40	2.35	2.93
	Code B	2.63	3.10	2.17	2.74
Experiment 4, integrated encoding	Code A	2.80	3.54	2.43	3.09
	Code B	2.73	3.27	2.34	2.92

Note. Cell values are average observed d' values for the experimental conditions listed (by rows) given separately for participant samples and ideal-detector d' levels assessed (by columns). Signal beam d' values equaling 3.81 and 4.60 correspond to individual display-bin d' levels .241 and .291, respectively. The University of California, San Diego (UCSD), sample did not participate in the Experiment 1 baseline study. DOD = Department of Defense; TSD = theory of signal detection.

(cf. Ashby & Townsend, 1986; Garner, 1974; Maddox, 1992). The applied issue is the impact on luminance-code performance introduced by a noninformative source producing masking chromatic noise.

DOD and UCSD participants were separately assessed and might be expected to perform differently based on motivation factors alone, although observing similar performance trends across groups would lend confidence to the practical applicability of the tested coding systems. Experiment 2 may be expected to find performance differences arising from the two color codes. For example, comparing the chromatically saturated Code A with the uniformly desaturated Code B, we might expect to observe attenuated detection of a luminance code due to Code A's relatively higher chromatic masking in spite of the small spatial extent of the stimulus information elements (although such interference and masking of a luminance code is presumed to be less likely for targets smaller than 2° visual angle). However, Code B's on-average higher luminance (twice Code A's average luminance) might be expected to produce improved performance compared with Code A.

Method

Participants. Experiment 2 separately assessed 5 DOD participants and 18 UCSD participants.

Materials. Experiment 2's stimuli were identical to those presented in Experiment 1 with the following exceptions: (a) As schematically depicted in Figure 3b, in addition to the meaningful luminance signal embedded in luminance noise, the stimulus panels incorporated irrelevant chromatic noise in the displays; (b) two different color coding schemes were tested; and (c) only two d' levels were assessed. The rationale for these modifications were (a) the "color irrelevant" beam-detection performance would indicate the performance differential attributable to the introduction of nonmeaningful chromatic content to the detection task; (b) the assessment of two color codes may yield insights into the impact of perceptual factors on color-code efficacy through the performance differences arising from the two color codes' different properties; and (c) only two d' signal levels were tested because using two coding schemes doubled the number of trials for the 188 panels assessed, resulting in a total of 376 beam detection trials for the entire experiment (thus equaling the number of trials in Experiment 1). Other differences between the UCSD and DOD assessments are detailed above in the *General Method* section.

Figure 1b shows a typical Experiment 2 stimulus, and response format is shown in Figure 1d. Irrelevant chromatic noise was applied to the Experiment 2 stimuli in a manner that paralleled the monochromatic noise indices described above. All 188 stimuli were colored once using Code A and once using Code B.

Design. All trials in Experiment 2 were 94-alternative forced-choice tasks.

Procedure. As in Experiment 1, in Experiment 2 participants located the single signal beam (i.e., the "brightest beam"). All other aspects of the procedure for Experiment 2 were identical to Experiment 1.

Results

Given Experiment 1's baseline detection performance, the goal of Experiments 2 was to determine whether the introduction of random, or *irrelevant,* color would impact detection performance in our displays. Existing work on similar images is equivocal on whether added chromatic contrast will impact, negatively or positively, detection performance. Three issues were addressed by Experiment 2: (a) Is the DOD detection performance similar to that seen in Experiment 1? (b) Do the two participant groups (DOD vs. UCSD) perform similarly? And (c) do both color codes yield similar detection performance?

Table 2, rows 2 and 3, presents the average d'_O values for each sample. Both groups detected signals for luminance-encoded data masked by chromatic noise at levels comparable with those found in Experiment 1. For DOD participants who took part in both Experiments 1 and 2, the difference in detection performance between the average monochrome-encoding performance versus Code A luminance encoding was nonsignificant (p = .13). The UCSD participants' performance differed from the Experiment 1 DOD performance (p = .04, two-tailed, using a t test for unequal samples), although their performance trends paralleled those of the DOD participants and are consistent with expected differences in expertise and motivation between the groups. Performance under both Code A and Code B show significant monotonic increases in performance with increasing tested d' levels for both participant groups ($p < .01$). This result suggests the underlying detection model is appropriately tracking observed detection performance. Considering data from both Code A and Code B, we did not find a significant difference in the performance of DOD participants compared with UCSD participants ($p = .072$, two-tailed, using a t test for unequal samples). Finally, considering data from both participant groups and collapsing across the two d' levels, we found that detection performance under Code A was significantly worse than that under Code B ($p < .01$). Comparing separately for each participant group, we found that UCSD participants' Code A performance was also significantly worse than that for Code B ($p < .01$), whereas the same comparison for the DOD sample shows no significant difference in performance using either Code A or Code B ($p = .540$).

Discussion

Experiment 2 shows that for the DOD sample, adding nonmeaningful chromatic content does not detract from Experiment 1's baseline detection performance. However, if we compare UCSD participants' Experiment 2 performance to the DOD Experiment 1 baseline, then we observe a performance difference that is most likely due to group-performance differences rather than encoding differences between Experiments 1 and 2. The finding that UCSD and DOD participants perform in a manner consistent with known population differences, and at levels similar to baseline, suggests that the addition of nonmeaningful chromatic content need not detract from detection performance. These results run counter to some data reviewed by Christ (1975) that suggest that color can detract from search performance even when it is used only to redundantly—or irrelevantly—encode information and do not support the idea that probabilistic independence produces inhibition across the luminance and chromatic channels. They do, however, support Garner's (1974) suggestion that hue and brightness are separable stimulus attributes that can be attended to without interference (Burns & Shepp, 1988). Overall, Code B produced significantly better detection performance compared with Code A. This is not unexpected given that Code B has lower and uniform average chromatic contrast and greater average luminance compared with Code A. These results are discussed further in the General Discussion section.

Experiment 3: Chromaticity-Encoded Signals Masked by Luminance Noise

Experiment 3 further examined the brightness–color information-processing relationship assessed in Experiment 2. In applied passive-sonar situations, operators typically (a) detect signal beams, then (b) classify detected beams. Thus, Experiment 3 assessed participants' ability to use chromatic information to classify color-coded signals masked by luminance noise into semantic categories. Semantic categories possibly denoted by the present coding system are reddish for a dangerous signal and greenish for a safe signal.

Note that the information-coding relationship used in Experiment 2 described above was reversed in Experiment 3, in which signal information was encoded by chromaticity, and both chromaticity and luminance encoded masking noise. The perceptual consequences are that signal and noise beams should, on average, look equally bright, but signal beams should, on average, look more red than noise beams. The coding of sensor data underlying this condition is schematically depicted in Figure 3c.

Experiment 3 is, in essence, the logical counterpart of Experiment 2's beam detection task in that the perceived brightness of any given beam carries no meaningful information, with brightness being randomly assigned to all the stimulus information units. Thus, the Experiment 3 classification data are driven strictly by the chromatic content of the stimuli.

Method

Participants. Experiment 3 assessed the same 5 DOD participants and 18 UCSD participants as in earlier experiments.

Materials. Luminance and chromaticity indices for noise display bins in Experiment 3 were identical to those noise display bins constructed for Experiment 2. New luminance and chromaticity indices for signal display bins were produced independently. First, a random value drawn from the noise distribution was quantized to obtain a color code's luminance index value. Next, the sum of a second random number and a d' value was quantized to obtain a code chromaticity index value. Experiment 3 incorporated design modifications based on the applied concerns that (a) the "luminance-irrelevant" beam classification performance will indicate the degree to which participants can use chromatic information in the presence of luminance noise when asked to semantically classify color-coded signals and (b) assessment of the two color-coding schemes in this manner will reveal performance differences that might arise from the two different color codes. That is, the highly saturated Code A might serve as a better color code. As in Experiment 2, the two color codes were tested in 94 different beam locations at two d' signal levels. Figure 1c shows an Experiment 3 stimulus constructed using Code B. An additional 188 images were constructed using Code A.

Design. Participants performed a yes–no signal classification task.

Procedure. Experimental trials used Code A and Code B stimuli, and participants classified a signal beam indicated by displayed cursors as either a "Reddish Signal" if the beam was more "Red" than "Neutral" and responded "Neutral Signal" otherwise. Beam classification instructions were not standardized for the DOD sample because participants were nonnaive researchers acquainted with the task. For naive UCSD participants the classification instructions specified the code-specific criteria for beam classification (all instructions are available on request). After studying the pre-indicated beam and selecting a response option by mouse click, participants could modify their classification if desired. Once satisfied with the classification, the "done" button was selected, and participants were immediately given feedback concerning the correct classification of the indicated signal beam (yellow cursor arrows or red cursor arrows indicated neutral and reddish signals, respectively). Feedback did not improve performance beyond initial training. All other test conditions were controlled as in Experiment 2.

Results

Recall that the chromatic signal information in Experiment 3 is masked by luminance noise and that the lowest level of luminance in Code A is black. The perceptual consequence of this is that a human observer cannot perceive any chromatic distinctions between display bins at the lowest level of luminance. Thus the chromatic d' is actually less than the statistical d' calculated from chromaticity indices. On average, because luminance is uniformly distributed, one eighth of the display bins will be black and therefore carry no chromatic information. We therefore adjusted the input beam d' from $\sqrt{250}$ to $\sqrt{0.875 * 250} = 14.79$. Thus, the two beam d' values reported in Table 2 (i.e., 3.81 and 4.60) were corrected to 3.56 and 4.3, respectively, for Experiment 3's Code A chromaticity-encoded signals.

The Gaussian ideal-detector model permits using a single observed hit and false-alarm pair to derive a unique d'_o value, thereby recovering d'_o from the hit and false-alarm rates of each participant. In addition, each d'_o derived from the hit and false-alarm data can be used to derive a corresponding proportion correct in a 94-alternative forced-choice test (Green & Swets, 1967, pp. 45–50).

Analysis issues were as follows: (a) Is performance similar to that of Experiment 2? (b) Do the two participant groups (DOD vs. UCSD) perform similarly? And (c) do the two color codes yield similar classification performance?

Table 2, rows 4 and 5, presents average d'_o values for each sample assessed. For both groups, when signal information was encoded by chromaticity, performance was inferior to Experiment 1's baseline.[5] Both groups also showed significant monotonic increases in performance with increasing d' levels ($p < .05$), excepting the nonsignificant difference between the two d' levels under Code B for the DOD sample ($p = .26$). These results suggest that the underlying detection model is appropriately tracking observed performance in the 94-alternative forced-choice hue-encoded task. Combining Code A and Code B data, we found that on average DOD participants performed significantly better than the UCSD group ($p < .01$, two-tailed, using a t test for unequal samples). Finally, considering both participant groups' data and collapsing over d' levels, we found that detection performance under Code A was significantly better than under Code B ($p < .01$). The same comparison made separately for each participant group showed that performance in each group under Code A was also significantly better than that for Code B ($p < .01$).

Discussion

In Experiment 3 participants effectively used a meaningful color code to classify signal beams into arbitrary classification categories. This was done in the absence of any meaningful luminance information that might complicate the ability to make use of chromatic information and suggests that in this display format color serves as an independent coding dimension and provides an ordered category structure. These results on ordered classification are new and promising for applied coding systems that aim to use color as a nonredundant salient coding dimension for ordered information categories. As in Experiment 2, Experiment 3 also shows that UCSD participants were not as skilled at the classification task as were the DOD participants. Again, this is not surprising given that the populations compared were university students and highly motivated and informed DOD participants. Finally, results suggest that Code A provides superior chromatic contrast compared with Code B. This is consistent with the fact that Code B's construction process yielded noticeably less saturated colors as a result of the contribution of blue phosphors. These results are discussed further below.

Experiment 4: Luminance–Chromaticity Integration

Having quantified in Experiments 2 and 3 performances with luminance information masked by chromaticity noise and chromaticity information masked by luminance noise, our next objective was to quantify performance with simultaneous luminance- and chromaticity-encoded information. As mentioned earlier, it could be the case that in the latter situation, in which two types of signal information are present, interference (or orthogonality loss) might occur, resulting in an overall performance that is worse than when encoding by luminance or chromaticity alone (Garner, 1974). Although this kind of inhibition is less common than summation across information channels, it can produce small decrements in performance in tasks using disparate information from multiple channels, compared with tasks in which all information derives from a single channel (see Graham, 1989, pp. 471–472, for a discussion). The purpose of Experiment 4 was to quantify any processing losses associated with simultaneously presented independent luminance and chromaticity encoding.

Experiment 4's stimulus panels used both meaningful luminance information and meaningful chromaticity information to encode the underlying sensor data (see Figure 1d). This modification permitted assessment of a combined brightness and hue information code capable of simultaneously conveying two different kinds of information with a single code value.

[5] What is the basis for comparing results from yes–no and 94-alternative forced-choice tasks? For a theoretical ideal observer, yes–no and forced-choice responses are related in the underlying TSD decision model (Green & Swets, 1967, pp. 45–50). In addition, the two tasks are highly perceptual and have been shown to be accurately modeled by TSD (Hershman et al., 1987). Thus, it is reasonable to compare empirical performance in the two tasks. Green and Swets illustrated how the percentage correct in a two-alternative forced-choice task is simply the area under the yes–no receiver-operating characteristic (ROC) curve (see their Figure 2-6, and pp. 47–48). They further explained how "the argument is easily extended to give the percentage correct in m-alternative forced choice designs. . . . thus detection theory provides a means of predicting the percentage of correct detections in forced choice from the yes–no or rating ROC curve" (p. 47). In practice, comparing d'_o values between yes–no, two-alternative forced-choice, and N-alternative forced-choice perceptual tasks (involving identical stimuli and participants) is common in sensory perception research and typically produces consistent d'_o values. Such comparisons under more cognitive tasks can yield inconsistent d'_o values. Our comparisons of Experiment 3 d'_o values with d'_o values from Experiments 1, 2, and 4 are considered appropriate and reasonable because (a) the tasks are highly perceptual, (b) all evidence suggests that the statistics of the data underlying the stimuli fit the TSD assumptions, and (c) across experiments we used identical stimulus panels (differing only by encoding condition) and participants. Although the appropriateness of comparing d'_o values from yes–no and 94-alternative forced-choice tasks may remain an issue, the main point is not the absolute efficiency of luminance or chromaticity coding but the fact that observers exhibit behavior consistent with relatively independent processing in the brightness and hue channels and efficient integration of combined information. The present experiments chose the yes–no and 94-alternative forced-choice tasks to simulate the demands in applied detect and localize sonar situations. If we had chosen to exclusively use either yes–no or forced-choice designs in our study, we would not expect our experimental results to be different.

Method

Participants. Experiment 4 separately assessed the same 5 DOD participants and 18 UCSD participants previously described.

Materials. Stimulus images were created for Experiment 4 by constructing new signal display-bin indices for the images used in Experiment 2, similar to that described above (details of all quantization methods used are available on request).

Design. Except for the two-dimensional information code used, Experiment 4's design, task, and apparatuses were identical to those of Experiment 2.

Procedure. Only an appropriate modification of the instructions differentiates Experiment 4 from Experiment 2. That is, participants understood that they should select the "brightest and reddest" beam and that both brightness and redness features were needed to indicate a signal. Figure 1d exemplifies the stimuli used in Experiment 4 in which the signal is encoded by both luminance and chromaticity.

Results

If d'_L, d'_C, and d'_I denote the beam d's for luminance information, chromaticity information, and optimally integrated luminance and chromaticity information, then, analogous to Equation 1, d'_I is defined as follows:

$$d'_I = \sqrt{d_L'^2 + d'_C}. \quad (3)$$

Given the way the stimuli were constructed, for an ideal observer we should have $d'_C = \sqrt{0.875} * d'_L$ for Code A and $d'_C = d'_L$ for Code B. Table 3 shows the values of d'_L, d'_C, and d'_I for both Codes A and B. Thus, performance in Experiment 4 was predicted by assuming the observer will make optimal use of the information obtained from luminance and chromaticity (Equation 3). For participants who took part in both Experiments 2 and 3, the individual d'_o values obtained in Experiment 2 estimate an observer's sensitivity for the luminance signal, and Experiment 3 d'_o values estimate sensitivity for chromatic signal strength. Equation 3 predictions are given in Table 2, rows 6 and 7, and are compared below with observed performance.

Experiment 4 analyses addressed the following questions: (a) Do Experiment 4 results accord with performance predicted by the detection model? (b) Do participants (DOD vs. UCSD) perform in ways similar to those seen in previous experiments? And (c) do the two color codes yield similar detection performance? Experiment 4's results are of particular interest in view of the perceptual channel considerations discussed earlier. That is, compared with a luminance code alone, cross-channel interference and masking might be expected to produce comparatively depressed performance in the saturated Code A combined code. Alternatively, if luminance and chromaticity are integral components, then a saturated combined code may facilitate performance (Ashby & Townsend, 1986). The existing literature has yet to address these empirical issues for the present display format.

Table 2, rows 8 and 9, presents the average d'_o values for each sample assessed, which can be compared with the performance predicted by the TSD model using the Experiment 2 and Experiment 3 data.[6] For both participant groups and color codes, performance under the two-dimensional codes was superior to that seen in the Experiment 1 baseline, and it was also significantly better *than performance* predicted by the underlying model ($p < .01$). Expected significant monotonic increases were seen in performance with increasing d' levels for both participant groups ($p < .05$), with the exception of the nonsignificant difference between the d' levels under Code B for the DOD sample ($p = .69$). Considering data from both Code A and Code B, we found that on average DOD participants performed the task significantly better than the UCSD participant group ($p < .05$, two-tailed, using a t test for unequal samples). Finally, considering data from both participant groups and collapsing across d' levels, we found that detection performance under Code A was significantly better than that found under Code B ($p < .01$). If we make the same comparison separately for each participant group, we find that performance in each group under Code A is also significantly better than that for Code B ($p < .01$).

Discussion

Experiment 4 shows that detection performance was better than expected in both groups. Initially, this seems a surprising result; however, one plausible explanation may be that Experiment 3 d'_o values underestimate the chromatic d'_o values achievable by participants in Experiment 4. Recall that in Experiment 3 luminance indices were uniformly distributed over the entire stimulus, whereas in Experiment 4 the signal beam included both a luminance signal and a chromaticity signal. Thus, on average signal beams in Experiment 4 were brighter than signal beams in Experiment 3. Two related perceptual consequences may produce Experiment 4's superior chromatic information processing: First, Experiment 4's "improved" luminance information may make it easier for the observer to see which display bins should be integrated. Second, an increase in the average luminance may yield improved chromatic discrimination and detection. Decreases in chromatic contrast with lower luminance indices is obvious in both color tables. For example, in Figure 3's approximation of Code A the appearance of an Experiment 3 signal display bin tends to correspond to one of the code values in the left half of the table. By comparison, the color appearance of a signal display bin in Experiment 4 tends toward code values in the upper left quadrant, where color discrimination between code elements is noticeably better.

Experiment 4 results may also not be unexpected given existing results on d' *additivity* (Pelli, 1985), or *facilitation* or *negative masking* (Nachmias & Kocher, 1970; Nachmias & Sansbury, 1974). This concept implies that participants might detect and discriminate contrasts in the combined luminance–hue stimulus that differ less than the smallest contrast that can be detected in either the luminance or chromatic stimuli alone. That is, sensitivity is heightened for a two-dimensional code stimulus compared with a one-dimensional code stimulus. In addition, studies on probability summation effects (i.e., when a compound stimulus is more detectable than either of two component stimuli) might also suggest that when two components such as luminance and chroma-

[6] It should be noted that Experiment 1's baseline performance curve is expected by theory to reflect inferior sensitivity compared with that seen in the combined code condition because in Experiment 4 d' is defined by Equation 3 (see the relatively high predicted-combined curve represented by the solid black line in Figure 6). Experiment 1's results are included in Figure 6 to only serve as a performance "point-of-reference" and to be consistent with the earlier presentations of Experiments 2 and 3 results.

Table 3
Signal Beam d' Values for Code A and Code B for Luminance-, Chromatic-, and Combined-Code Conditions

Code A			Code B		
Luminance	Chromaticity	Combined	Luminance	Chromaticity	Combined
3.81	3.56	5.21	3.81	3.81	5.35
4.6	4.3	6.29	4.6	4.6	6.5

Note. Cell values given in columns are the two input beam d' values tested for each encoding condition assessed.

ticity combine in a code, the result is facilitated detection performance (e.g., Graham, 1977, 1989; Sachs, Nachmias, & Robson, 1971). And, as mentioned earlier, facilitation is also possible through gestalt processing of information on all integral dimensional stimuli (Garner, 1974). These possible explanations (underestimated chromatic d' and perceptual facilitation) of Experiment 4 performance cannot be disentangled in the present study. However, Experiment 4's results suggest that either our component d' values were not estimated properly by Equation 3 or that the standard integration formula may not be appropriate for estimating performance in the present task given that perceptual facilitation is possible. As suggested by Luce (1994), applications of TSD toward integrating multiple signal sources in a composite code may require further formal modeling. Similar to earlier experiments, Experiment 4 shows that the DOD participants performed better than the naive UCSD participants. Code A performance was found to be superior to that under Code B.

Comparing results across experiments, we present in Figure 6 within-sample performance across Experiments 2, 3, and 4, relative to Experiment 1's monochrome baseline. Figure 6a presents the mean d'_o comparisons (collapsed across color code) for all the DOD experiments, and Figure 6b presents the analogous data from the UCSD experiments. Figure 6 demonstrates similar patterns of performance across experiments for the DOD and UCSD samples. For the present display format, both samples' performance for detection of a brightness code in the presence of irrelevant noise (Experiment 2) was found to be better than performance for simply classifying color-encoded stimuli (Experiment 3). Relative to the monochrome baseline, chromatic masking did not degrade detection performance (Experiment 2). Moreover, as discussed above, for both samples, performance under a combined information code (Experiments 4) yielded performance superior to that predicted by theory (the predicted-combined curve).

The DOD results in Figure 6a can perhaps be interpreted by a generalization of Garner's concepts of orthogonality loss and redundancy gain (Garner, 1974, pp. 124–128; Garner & Morton, 1969). For example, the DOD group's performance in Experiment 2 (luminance-encoded signals in luminance noise with irrelevant chromatic noise) suggests that orthogonality loss does not occur when irrelevant color is added to Experiment 1 stimuli (luminance-encoded signals in luminance noise at constant chromaticity), in that the results from the two experiments are not significantly different. Similarly, the DOD group's performance in Experiment 4 seems to suggest that orthogonality loss is not impairing performance in these multidimensional stimuli and that redundancy gain may be partially responsible for the superior performance seen under Experiment 4's multidimensional-encoding condition when compared with conditions in which signal encoding was achieved by a single dimension (Experiments 2 and 3). Similar interpretations may apply for other observer-group data; however, further experiments are needed before orthogonality loss and redundancy gain can be made specific for general observer performance on multidimensional stimuli such as those used here.

Figure 7 summarizes, for all participants assessed, average d'_o measures on (a) stimuli with signals based on two independent sources of information simultaneously encoded with luminance and chromaticity (left data series), (b) luminance-encoded signals masked by uniform noise (middle data series), and (c) chromaticity-encoded signals masked by uniform noise (right data series). As can be seen, there is a substantial effect due to encoding method and a lesser effect due to color code.

General Discussion

Specifically, what do the present results suggest for applications of multidimensional information codes to gradient intensity data formats? (a) A luminance code can be presented in the context of nonmeaningful hue variation without negatively affecting the processing of a luminance signal, (b) a hue code can be used to independently convey information in the context of nonmeaningful luminance variation, (c) increased processing demands of a hue–luminance code do not produce decreases in performance, and (d) *in this display format TSD is reasonably useful as a statistical tool* for modeling and predicting observer performance in that observed d' was found to monotonically increase with increases in tested d'. However, TSD's integration formula underestimated sensitivity under a combined brightness–hue code (Experiment 4). Although this is an aspect of the formal model needing appropriate modification, it is perhaps not surprising given existing results that suggest facilitated detection for multichannel stimuli. Contrary to our expectations and to some suggestions in the literature, the introduction of chromatic noise does not substantially depress detection of a luminance signal compared with a monochromatic baseline (Experiment 2). Also somewhat unexpected is the finding that color codes can be effectively used in the presence of luminance noise (Experiment 3), although—not surprisingly given the small size of the information elements in our stimuli—not as effectively as a luminance code. Overall, these data show that processing chromaticity-encoded information masked by luminance noise is significantly more difficult than processing luminance-encoded information masked by chromatic noise.

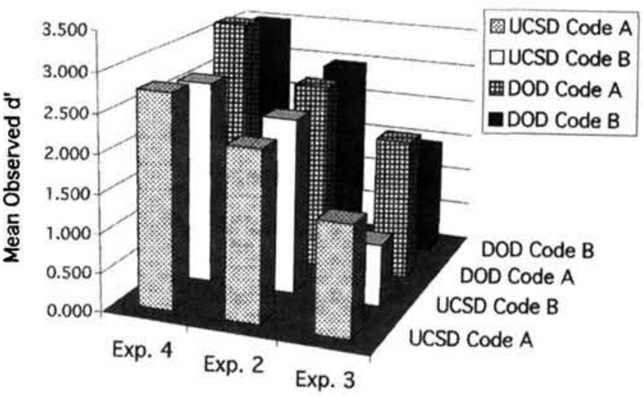

Figure 7. Performance data from Experiments 2, 3, and 4 for two participant groups (participants from the Department of Defense; DOD; and from the University of California, San Diego; UCSD) and for two color codes (Codes A and B). Experiment 2 signals were encoded using luminance only. Experiment 3 signals were encoded by chromaticity. Experiment 4 signals were encoded by both luminance and chromaticity. Speckled bars represent performance data for UCSD Code A. White bars represent performance data for UCSD Code B. Hatched bars represent performance data for DOD Code A. Black bars represent performance data for DOD Code B.

Observed performance differences under Code A and Code B are summarized as follows. Detection performance under the monochrome code (Experiment 1) resembled that under both Codes A and B (Experiment 2); however, for luminance-based encoding, performance under Code B (the uniformly desaturated code) was superior to that under Code A (Experiment 2). This makes sense because lightness and saturation are held to be two dimensions that seem to be processed together (Garner, 1974). In Experiment 3's hue-based classification task, performance under Code A was superior to performance under Code B. This seems reasonable given that Code A has relatively higher chromatic contrast. Finally, Experiment 4's detection performance under integrated Code A was superior to that under integrated Code B. Again, Code A's greater chromatic contrast seems to convey a more discriminable signal, presumably by using a greater color differential. Nevertheless, the integrated code condition is clearly a situation in which neither dimension (brightness or hue) inhibits processing of the conjoined code. Interestingly, according to Ashby and Townsend (1986), if two component dimensions are separable, then Experiment 4 performance would be expected to mimic that of Experiment 2. However, if the two components are integral, then, as was observed, improved performance would be predicted for Experiment 4. Although the specific relations between these perceptual dimensions cannot be decided through the present experiments, the present data do support the notion that the chromatic differences that exist between the two color codes do not differentially inhibit luminance detection performance to any appreciable degree (Experiment 2).

Together data from Experiments 2 and 3 indicate that performance with luminance-encoded signals is substantially better than that for chromaticity-encoded signals. That performance with chromaticity-encoded signals is not as good as that with *luminance*-encoded signals is not surprising for two reasons. First, the detection task requires observers to visually integrate perceived colors across exactly those display bins that make up each beam. However, luminance masking can reduce the ability of the observer to perceive edges defined by chromatic differences. Thus, luminance masking can cause signal display bins to be perceived as belonging to neighboring noise beams, and vice versa. Consequently, because spatial discrimination of different chromaticities is not as good as spatial discrimination of different luminances, chromatic information in noise display bins is more likely to get included with chromatic information in signal display bins. Second, luminance processing is expected to be less dependent on large field size compared with chromatic processing (Widdel & Post, 1992). Thus, small visual angle of information display bins may produce a greater reduction in an observer's ability to discriminate between display bins on the basis of color relative to similar discriminations based on brightness.

Finally, we found results contrary to the notion that integration of both luminance and chromatic information in a single code hinders detection performance because of cross-channel inhibition effects (Experiment 4). That is, the integrated information codes we tested produced performance better than that observed for either code alone, or, put differently, better than that possible using the best channel (i.e., luminance-channel performance). This is a positive indicator for display designers seeking to pack more information into user interfaces using two-dimensional information codes.

What Do the Results for the UCSD and DOD Participant Samples Reveal?

How do varying levels of observer "expertise" (defined by experience and motivation of the participant) impact detection and classification performance? The results imply that although there were performance differences between the two samples, in general these differences are what might be expected given the two populations. It seems reasonable to assert that the tested two-dimensional codes do not require any special sophistication to interpret and thus are viable coding alternatives for information displays used by both trained and inexperienced operators.

How Do We Interpret Results for the Multidimensional Information Codes?

We have shown that one can pack an additional dimension of information into a display by adding a hue dimension to a brightness dimension. Color in an information code need not significantly depress detection performance for a brightness signal. This is an important new finding for information representation because it suggests interface designers can effectively use composite information codes when an application requires that additional information be conveyed by a user interface. The implication is that these dimensions are perceptually and cognitively distinct enough to be judged independently in conjoined codes.

Although the two color codes generally produced very similar performances, Code A was substantially more effective at color coding information in the classification task. This result follows from the fact that Code A is highly saturated (deeper) whereas Code B is more desaturated (or pastel). Thus, the recommendation made for applications involving pattern recognition that integrates small, noise-masked information elements over larger signal areas is that when combined with a brightness code, classification codes should maximize chromatic information when color encodes salient classification information, and desaturated codes should be avoided when very small field sizes are used. These results are consistent with known perceptual effects on dimensional interactions, even though we used nonstandard and novel stimuli masked by statistical noise.

As has been suggested elsewhere (Jameson, 1997; Jameson & D'Andrade, 1997), the present results support the notion that a person can independently respond to brightness information in spite of the fact that he or she is also processing color, which further supports the notion that the dimensions of hue, brightness, and saturation are perceptually salient and psychologically meaningful characteristics of human color processing. This is indicated by the findings that the dimensional constructs of brightness and hue are easily untangled in the present tasks, which have as their express aim the goal of pushing detection thresholds (using stringent d' levels). In general, our results suggest that multidimensional codes—at least two-dimensional ones—are viable, especially when the two dimensions are robust psychologically as is the case with the brightness and hue dimensions used here. Principled procedures were given for developing information codes that are extendable, that are generalizable across devices, and that address perceptual inconsistencies (e.g., hue percepts that vary with brightness contrast). We showed that such issues can be accounted for by an information code that uses a perceptual space as a basis for construction, without reducing information representation capabilities and without any significant effort beyond that normally invested when engineering-based methods are used to devise a physically based information code. On the basis of these results, we suggest the development and use of psychologically based codes for multidimensional information representation because they are feasible and effective.

References

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93*, 154–179.

Beer, R. D., Jameson, K. A., & Kaiwi, J. L. (1997, May). *Empirical assessment of detection and classification performance with a two-dimensional color-code.* Abstract (No. 5480) presented at the 1997 Association for Research in Vision and Ophthalmology (ARVO) Conference, Ft. Lauderdale, FL.

Brainard, D. H. (1989). Calibration of a computer controlled color monitor. *COLOR Research and Application, 14*, 23–34.

Burns, B., & Shepp, B. E. (1988). Dimensional interactions and the structure of psychological space: The representation of hue, saturation and brightness. *Perception and Psychophysics, 43*, 494–507.

Carter, R. C., & Carter, E. C. (1988). Color coding for rapid location of small symbols. *COLOR Research and Application, 13*, 226–234.

Christ, R. E. (1975). Review and analysis of color coding research for visual displays. *Human Factors, 17*, 542–570.

Commission Internationale de l'Eclairage. (1931). *Proceedings, International Congress on Illumination.* Cambridge, England: Cambridge University Press.

Donderi, D. C. (1994). Visual acuity, color vision, and visual search performance at sea. *Human Factors, 36*(1), 129–144.

Garner, W. R. (1974). *The processing of information and structure.* Hillsdale, NJ: Erlbaum.

Garner, W. R., & Morton, J. (1969). Perceptual independence: Definitions, models, and experimental paradigms. *Psychological Bulletin, 72*, 233–259.

Graham, N. (1977). Visual detection of a periodic stimuli by probability summation among narrow band channels. *Vision Research, 17*, 637–652.

Graham, N. (1985). Detection and identification of near-threshold visual patterns. *Journal of the Optical Society of America, 2*, 1468–1481.

Graham, N. (1989). *Visual pattern analyzers.* New York: Oxford University Press.

Green, D. M., & Swets, J. A. (1967). *Signal detection theory and psychophysics.* New York: Wiley.

Hershman, R. L., Kaiwi, J. L., & Pilmore, E. (1987). *Tests of display sensitivity* (Tech. Note No. 87-15). San Diego, CA: Navy Personnel Research and Development Center.

Ishihara, S. (1994). *Ishihara's tests for colour-blindness* (concise edition). Tokyo, Japan: Kanehara & Co., Ltd.

Jameson, K., & D'Andrade, R. G. (1997). It's not really red, green, yellow, blue: An inquiry into perceptual color space. In C. L. Hardin & L. Maffi (Eds.), *Color categories in thought and language* (pp. 295–319). Cambridge, England: Cambridge University Press.

Jameson, K. A. (1997). What Saunders and van Brakel chose to ignore in color and cognition research. *Behavioral and Brain Sciences, 20*, 195–196.

Kadlec, H., & Townsend, J. T. (1992). Perceptual and decisional separability. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 181–227). Hillsdale, NJ: Erlbaum.

Kaiwi, J. L., Bamber, D., & Urban, C. (2000, June). *Color-coding digital information* (Tech. Rep. No. 1819). San Diego, CA: SPAWAR Systems Center.

Luce, R. D. (1994). Thurstone and sensory scaling: Then and now. *Psychological Review, 101*, 271–277.

Maddox, W. T. (1992). Perceptual and decisional separability. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 147–180). Hillsdale, NJ: Erlbaum.

Nachmias, J., & Kocher, E. C. (1970). Visual detection and discrimination of luminous increments. *Journal of the Optical Society of America, 64*, 1321–1327.

Nachmias, J., & Sansbury, R. V. (1974). Gating contrast: Discrimination may be better than detection. *Vision Research, 21*, 1041–1052.

Nagy, A. L., & Sanchez, R. R. (1990). Critical color difference determined with a visual search task. *Journal of the Optical Society of America, A7*, 1209–1217.

Nagy, A. L., & Sanchez, R. R. (1992). Chromaticity and luminance as coding dimensions in visual search. *Human Factors, 34*, 601–614.

Newhall, S. M., Nickerson, D., & Judd, D. (1943). Final report of the O.S.A. subcommittee on the spacing of the Munsell colors. *Journal of the Optical Society of America, 33*, 385–418.

Pelli, D. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America, 2*, 1508–1531.

Sachs, M. B., Nachmias, J., & Robson, J. G. (1971). Spatial frequency channels in human vision. *Journal of the Optical Society of America, 61*, 1176–1186.

Smallman, H. S., & Boynton, R. M. (1990). Segregation of basic colors in an information display. *Journal of the Optical Society of America, A7*, 1985–1994.

Smallman, H. S., & Boynton, R. M. (1993). On the usefulness of basic colour coding in an information display. *Displays, 14*(3), 158–165.

Thomas, J. P. (1985). Effect of static-noise and grating masks on detection and identification of grating targets. *Journal of the Optical Society of America, 2,* 1586–1592.

Tversky, A., & Krantz, D. H. (1970). The dimensional representation and the metric structure of similarity data. *Journal of Mathematical Psychology, 7,* 572–597.

Van Orden, K., & Benoit, S. (1993). Color recommendations for prototype maritime tactical displays (Tech. Rep. No. AD B184953 [1192]). Groton, CT: Naval Submarine Medical Research Laboratory.

Widdel, H., & Post, D. L. (1992). *Color in electronic displays.* New York: Plenum.

Wyszecki, G., & Stiles, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulae.* New York: Wiley.

Received August 1, 1997
Revision received January 3, 2001
Accepted January 19, 2001 ■