

1 **Rapid measurement and machine learning classification of color vision** 2 **deficiency**

3
4 Jingyi He¹: j.he@northeastern.edu

5 Peter J. Bex^{1*}: p.bex@northeastern.edu

6 Jan Skerswetat¹: j.skerswetat@northeastern.edu

7 ¹Department of Psychology, Northeastern University, USA

8 *Corresponding author

9 10 11 **Abstract**

12 Color vision deficiencies (CVDs) indicate potential genetic variations and can be important
13 biomarkers of acquired impairment in many neuro-ophthalmic diseases. However, CVDs are
14 typically measured with insensitive or inefficient tools that are designed to classify dichromacy
15 subtypes rather than track changes in sensitivity. We introduce FInD (Foraging Interactive D-
16 prime), a novel computer-based, generalizable, rapid, self-administered vision assessment tool and
17 applied it to color vision testing. This signal detection theory-based adaptive paradigm computes
18 test stimulus intensity from d-prime analysis. Stimuli were chromatic gaussian blobs in dynamic
19 luminance noise, and participants clicked on cells that contain chromatic blobs (detection) or blob
20 pairs of differing colors (discrimination). Sensitivity and repeatability of FInD Color tasks were
21 compared against HRR, FM100 hue tests in 19 color-normal and 18 color-atypical, age-matched
22 observers. Rayleigh color match was completed as well. Detection and Discrimination thresholds
23 were higher for atypical observers than for typical observers, with selective threshold elevations
24 corresponding to unique CVD types. Classifications of CVD type and severity via unsupervised
25 machine learning confirmed functional subtypes. FInD tasks reliably detect CVD and may serve
26 as valuable tools in basic and clinical color vision science.

27
28
29 **Keywords:** color detection, color discrimination, color vision deficiency, colorblindness, cone-
30 isolating directions, unsupervised machine learning, K-means clustering, vision diagnostics

31 32 33 **Introduction**

34 Conventional phenotypical categories for inherited color vision deficiency (CVD) are anomalous
35 trichromats (AT), dichromats, and monochromacy, with mild to strong color vision defects,
36 respectively. They can be further referred to as protan, deutan, or tritan types, with L-, M-, and S-
37 cone relevant deficiencies, respectively(1). Identification and diagnosis of CVD are critical in
38 many respects. In clinical applications, abnormal color vision may reveal genetic dyschromatopsia
39 or early signs of neural pathway or systemic diseases(2). However, widely employed color vision
40 tests are not ideal for detecting and monitoring acquired CVD progression or remediation, owing
41 to their insensitivity, long testing time, difficulty of administration, and interpretation of their
42 results.

43
44
45 Color matches conducted with an anomaloscope have been considered a gold standard for precise
46 CVD diagnosis. An anomaloscope enables red-green Rayleigh matching and blue-green Moreland

1 matching tests, in which the observer adjusts the mixed light side of the bipartite field to match the
2 light in the reference side(3). However, anomaloscopes are expensive and a full examination of
3 the color matches requires extensive instruction, expert administration, and exhaustive testing time.
4 In addition, matching ranges of an extreme anomalous trichromat can be indistinguishable from
5 those of dichromats. Pseudoisochromatic plates are capable of rapid screening and classification
6 of CVD. Ishihara plates contain pseudoisochromatic numbers and curved lines, as well as
7 vanishing plates with stimuli seen by only CVD patients. Hardy-Rand-Rittler (HRR) plates(4)
8 exceed Ishihara plates, in that they contain universal color symbols instead of Arabic numbers,
9 have a three-step severity scale for each CVD subtype, use a two-step psychometric protocol (i.e.,
10 what symbol do you see in which quadrant) to reduce probability of guessing, and are able to
11 classify tritan(5). However, printed pseudoisochromatic plates have a number of shortcomings.
12 They require a trained clinician to administer the test, and are generally insensitive to detect subtle
13 changes of color detection due to the absence of a severity scale (Ishihara), a coarse severity scale
14 (HRR), and the lack of cone-isolated colors. The Farnsworth-Munsell 100 hue test (FM100)
15 provides a relatively complete color discrimination measurement for Protan, Deutan, and Tritan,
16 by asking participants to arrange 85 caps according to color, but the task is extremely time-
17 consuming, as is the analysis of FM100 error scores. These issues are addressed by D-15, an
18 abridged version of FM100, however, interpretation of the error scores of both tests can be vague.
19 CVD patients can also rely on luminance cues to pass the arrangement task or improve test
20 scores(6).

21
22 Attempts to improve performance of traditional color vision tests have been made with computer-
23 based tests. The Cambridge Color Test (CCT) adopted the pseudoisochromatic pattern from the
24 above-described printed tests with a colored Landolt C orientation-identification stimulus
25 embedded in static luminance noise, and utilizes forced-choice adaptive procedures(7). The Color
26 Assessment and Diagnosis (CAD) test employs a moving chromatic square superimposed onto
27 dynamic luminance noise(8). Resulting thresholds of both tests in vector length units can be used
28 to classify the type and severity of CVD. In clinical contexts, testing duration is crucial. The CCT
29 Trivector test shortened the testing procedure by assessing only the three confusion lines. Faster
30 and simpler tablet-based tests which measure detectability along the three confusion lines were
31 developed for better accessibility and for young children(9). The trade-off between testing speed
32 and information collected has been the main cause of the disadvantages in these color vision tests.

33
34 In this article, we introduce and validate a generalizable, rapid, and self-administered computer-
35 based procedure named FInD (Foraging Interactive D-prime), and adapt it to assess color
36 detectability and discriminability. FInD uses an adaptive algorithm that measures visual
37 performance thresholds for a wide range of visual functions, e.g. contrast sensitivity, motion- and
38 form-coherence(10, 11). The algorithm selects stimulus strength derived from d' , a signal-to-noise
39 ratio metric referring to the detectability of the stimulus, and calculates a new d' based on previous
40 responses.

41
42 FInD Color Detection measures detection thresholds to L, M and S cone-isolating stimuli,
43 reflecting the performance of individual cone types, while FInD Color Discrimination measures
44 hue discrimination thresholds around 6 directions on a HSV (Hue, Saturation, Value) or
45 equiluminant color plane. The two color spaces are used for different purposes – while
46 equiluminant colors satisfy accurate measurement of color discrimination, HSV colors do not

1 require careful calibrations so can be readily applied. The detection and discrimination tasks
2 together interrogate an observer's color vision system: the detection task classifies photoreceptor-
3 level color sensitivity, and the discrimination task further quantifies the resolution of color
4 perception. The computer-based Rabin Cone Contrast Test (RCCT) also adopted cone-isolating
5 colors to screen CVD(12), in which color letters were presented without any background noise.
6 FInD Color tasks instead use Gaussian blobs, therefore remove high frequency chromatic signals
7 and avoid optotype familiarity. Dynamic luminance contrast noise is also added to mask potential
8 luminance artifacts. Our first aim was to validate FInD Color detection and discrimination in terms
9 of thresholds, testing duration and reliability in both color-normal (CN) and CVD groups. The
10 second aim was to determine CVD type and severity using unsupervised machine learning (UML)
11 classification with FInD Color detection and discrimination thresholds. The classification of CVD
12 subtypes is difficult because genetic testing is expensive and laborious, and results of current
13 behavioral tests for thresholds vary due to different stimulus types and tasks. Personalized
14 threshold results generated with FInD enable the deployment of UML approaches to determine
15 groups based purely on the behavioral performance of detectability and discriminability. The third
16 aim was to compare FInD results against that of conventional clinical tools— HRR and FM100.

17 18 **Methods**

19 The experiment protocol was approved by Northeastern University Institutional Review Board and
20 followed the principles in the Declaration of Helsinki.

21 22 ***Participants***

23 19 participants (mean \pm SD age: 26.2 ± 10.0 ; age range: 18-50; 9 females) with self-reported
24 normal color vision and 18 (mean \pm SD age: 23.1 ± 7.8 ; age range: 18-54; 1 female) with self-
25 reported inherited color vision deficiency were recruited after providing informed consent and
26 completing a demography and ocular history questionnaire. All participants had normal (20/20 or
27 better) or corrected-to-normal visual acuity and had no history of eye diseases except for one
28 participant with strabismus and two participants with amblyopia, among whom one amblyopia
29 participant also had defective color vision. Two participants were diagnosed with attention deficit
30 disorder and depression, respectively, and both received medical treatments.

31 32 ***Apparatus***

33 Experimental procedures were programmed by Psychtoolbox(13) in MATLAB (MathWorks,
34 Natick, MA), and presented on a 32" 4K LG display with a resolution of 3840 \times 2160. The display
35 was gamma-corrected with SpyderX elite colorimeter (Datacolor, Lawrenceville, NJ) and the
36 spectra and luminance were measured with a Photo Research PR-650 spectroradiometer (Photo
37 Research, Chatsworth, CA). Luminance of the mid-grey background was 90.3 cd/m². Participants
38 viewed the screen binocularly at a distance of 111 cm, subtending a visual angle of 35 deg \times 20
39 deg, with the head position stabilized in a chin rest. Standard illumination applied for HRR and
40 FM100 administration was provided by a Sol•Source daylight lamp (117V, 50/60 Hz)
41 manufactured by GretagMacbeth. The time was recorded using a standard mobile phone timer
42 application or by the computer for FInD tests.

43 44 ***Tasks and Stimuli***

1 Four methods were compared in this study: Hardy-Rand-Rittler (HRR) Pseudoisochromatic Plates
2 (4th Edition), Farnsworth-Munsell 100 hue test (FM100), Rayleigh color match, and FInD Color
3 Detection and Discrimination tasks.

4
5 *Hardy-Rand-Rittler (HRR) Pseudoisochromatic Plates*

6 HRR was conducted by the experimenter flipping the plates and participants reporting the shape
7 and location of test color symbols under controlled lighting.

8
9 *Farnsworth-Munsell 100 hue test (FM100)*

10 Participants were asked to arrange the 85 caps according to reference colors in each testing case.
11 The four cases were completed in random order.

12
13 *Rayleigh Color Match*

14 Rayleigh color matching data were collected on 14/19 CN and 6/18 CVD with an Oculus HMC
15 Anomaloscope (Oculus, Germany). Participants were asked to complete 8 measurements (4 with
16 each eye) and an additional matching range identification trial.

17
18 *FInD Color detection*

19 Stimuli utilized in FInD Color Detection task were cone-isolated Gaussian blobs ($\sigma=1^\circ$, support
20 diameter = 4°) (Figure 4a left). L-, M-, and S-cone isolating directions in RGB unit were calculated
21 by integrating Stockman-Sharpe cone fundamentals(14, 15) and the measured display spectra, then
22 weighted by cone contrasts (detailed computation steps can be found in He, Taveras-Cruz (16)
23 Appendix; see Supplementary materials for exact values used). Stimulus contrast was adaptively
24 controlled by the FInD algorithm.

25
26 *FInD Color discrimination*

27 FInD Color Discrimination task measures discriminability of a pair of two small Gaussian blobs
28 ($\sigma=0.6^\circ$, support diameter= 3.6°) with different colors (Figure 4a right). Each stimulus contains a
29 pair of colors that were selected from the HSV color space (Figure 4b). Six hues (H = red, yellow,
30 green, cyan, blue, magenta), 3 primary and 3 confusion axes, in the half-brightness ($V=0.5$) HSV
31 plane were tested separately with two saturation levels ($S = 100\%$ or 50%). For each selected hue
32 axis, two stimulus colors were selected at the same angular distance away from this hue axis in
33 opposite directions. The angular distance between test colors in HSV space was adaptively
34 controlled by the FInD algorithm.

35
36 ***Procedures***

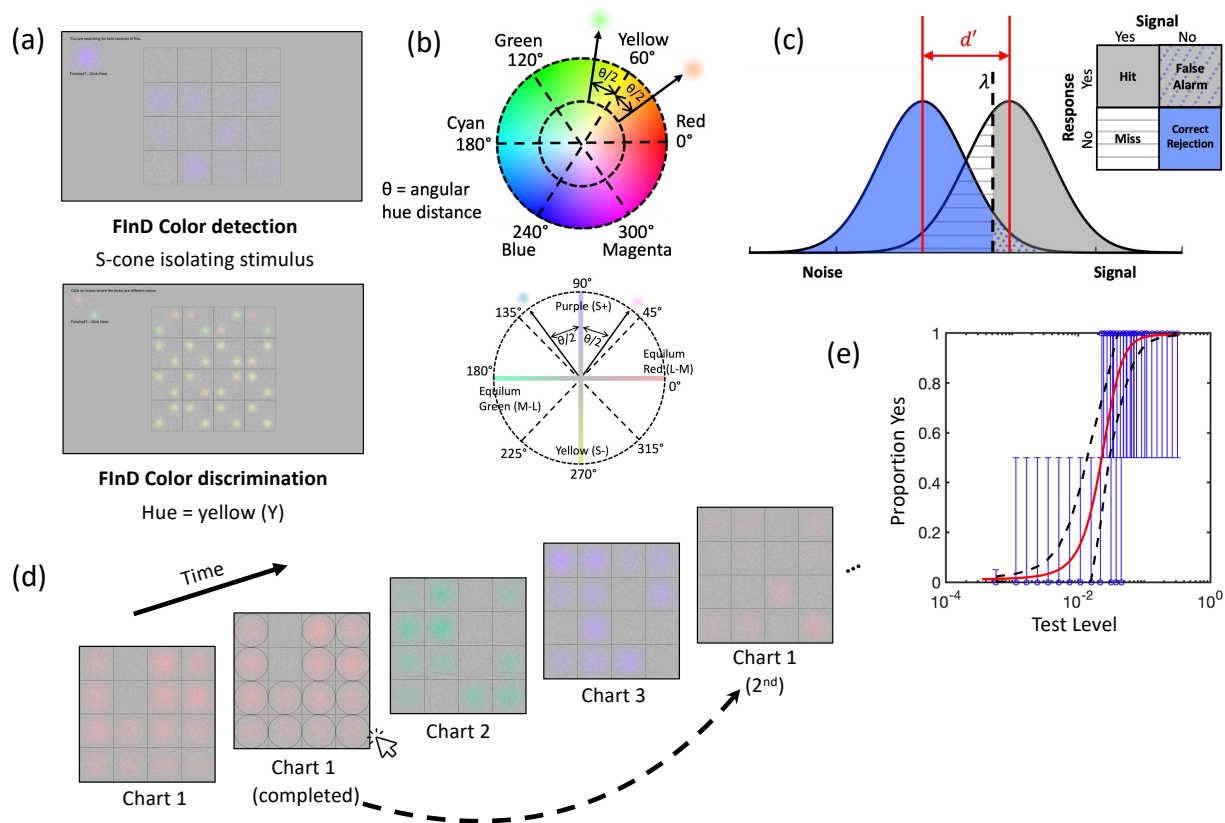
37 Each participant completed the four tasks once (test session) or twice (test-retest sessions) in
38 random order after optometric screening.

39
40 The FInD adaptive algorithm measures d' to efficiently estimate thresholds. In short, d' is a
41 measure of detectability or discriminability in signal detection theory. As in Figure 4c, d' is the
42 distance between the noise and signal distribution means, and the location of the criterion (λ)
43 directly affects the proportion of responses corresponding to “hit”, “miss”, “false alarm”, and
44 “correct rejection”. The initial stimulus contrast (detection) or angle distance (discrimination) was
45 determined by pre-defined approximate estimates of slope and threshold to derive the test range to
46 span difficult ($d'=0.1$) to easy ($d'=4.5$) stimuli. The independent variable of the stimuli for the first

1 chart were sampled in log steps across this test range. Subsequent charts use the combined
 2 responses across all previous charts to re-estimate d' (Figure 4e).

3
 4 Stimuli were presented in charts (Figure 4a) containing 4×4 cells ($6^\circ \times 6^\circ$) at the center of the screen
 5 with each cell containing one stimulus at pre-selected contrasts (detection) or hue difference
 6 (discrimination) embedded in 8 Hz dynamic luminance noise, with check size 10 arcmin and $\pm 20\%$
 7 luminance contrast. A high stimulus-intensity example and instructions were provided at the upper
 8 left corner of the screen to help participants identify the targets to be foraged. Participants were
 9 informed that target stimuli were present in some but not all cells, and the number of targets varied
 10 from chart to chart. They were instructed to click on any cells that contained “faint versions” of
 11 the stimulus (detection) or on cells that contained blob pairs that have different colors
 12 (discrimination). The chart stayed on the screen until the observer completed the current chart by
 13 clicking on the exemplar. d' was then calculated based on hit/ miss/ correct rejection/ false alarm
 14 classifications as a function of stimulus contrast (detection) or hue difference (discrimination). The
 15 estimate of d' from all previous responses was used to generate the stimulus range (contrast or hue
 16 difference) in the next chart. A series of three charts for each stimulus were presented in interleaved
 17 order (Figure 4d).

18



19
 20 Figures 4. Illustration of FInD stimuli and experimental procedures. (a) FInD detection (top) and
 21 discrimination (bottom) task interfaces. (b) The top color wheel shows a cross-section of the HSV
 22 space for $V=1$, from which six hues (0° to 300° in 60° steps) and two saturation levels (0.5 and 1)
 23 were chosen and used in the discrimination task. If, for instance, yellow (60°) discrimination were
 24 tested, two colors are symmetrically selected the same angular hue distance ($\theta/2$) away from
 25 yellow with a fixed saturation level. The bottom color wheel shows the equiluminance plane with

1 four primary axes representing the red-green and blue-yellow postreceptoral mechanisms (c)
2 Illustration of signal detection theory. The noise distribution (blue) and signal distribution (grey)
3 bell curves lie on the normalized Z-score abscissa. Detectability or discriminability (d') and
4 criterion (λ) are depicted. The areas under the curves correspond to “hit”, “miss”, “false alarm”,
5 and “correct rejection”, respectively, according to stimulus presentation and response. (d) FInD
6 experimental procedures with cone isolating direction detection stimuli as an example. The dashed
7 arrow represents the adaptive procedure that selects a range of stimulus intensities on the 2nd chart
8 based on analysis of the responses to stimuli on the 1st chart. (e) An example of a typical
9 psychometric function: blue data show the probability that the observer reported the presence of a
10 stimulus as a function of intensity, vertical lines show binomial standard deviation, red curve
11 shows the best fitting function for Eq.1, and black dashed lines represent upper and lower 95%
12 confidence intervals. The separate data point on the left of other data points indicates the false
13 alarm rate.

14

$$15 \quad \text{Proportion Yes} = 1 - \Phi \left(\Phi^{-1}(1 - \text{FalseAlarmRate}) - \frac{5 \times \left(\frac{\text{TestLevel}}{\text{Threshold}} \right)^{\text{slope}}}{\sqrt{(5^2 - 1) + \left(\frac{\text{TestLevel}}{\text{Threshold}} \right)^{2 \times \text{slope}}}} \right),$$

16 where Φ is standard normal cumulative distribution function. Eq.1

17
18

19 Results

20 Data analyses were performed in MATLAB (MathWorks, Natick, MA). Descriptive statistics of
21 FM100 and FInD tasks for the CN group are summarized in Table S1. The analyses below are
22 based only on the test session data except for the repeatability analysis.

23

24 HRR

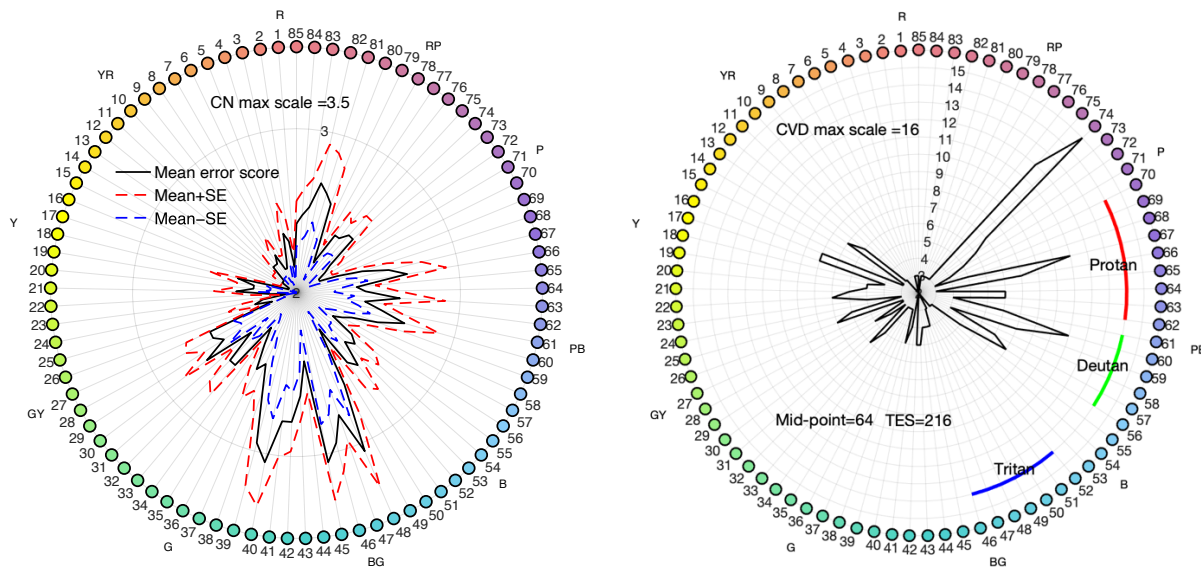
25 All CN observers passed the first ten diagnostic plates of HRR. The time recorded for 12 CNs to
26 pass was 37 seconds on average. All CVDs self-reported to have color vision deficits failed HRR
27 and were classified as protan or deutan with mild, medium, or strong defect (Table 1), and time
28 spent for 13 CVDs was 2'53" on average to complete all 24 plates of HRR. One CVD observer
29 with mild red-green deficiency achieved equal protan and deutan scores thus could not be classified.

30

31 FM100

32 Total error scores (TES) and right-half mid-point (MP) were reported for FM100. TES is
33 calculated as the sum of the error scores for each color with 2 subtracted from each error score.
34 For CNs, TES is expected to range from 0 to 100 where superior, average, and low discrimination
35 abilities are indicated by error scores ranging from 0~16, 20~100, and more than 100. In our
36 sample 42%, 53%, and 5% of the CN observers respectively fell in the three categories (Table 1).
37 Mean and standard deviation of the TES of the CN group are reported in Table S1, and average
38 error score pattern is shown in the left panel of Figure 1 for the CN group. Note the radial axis
39 range, where the center of the error score pattern is 2, indicating perfect responses, with the largest
40 error score scale being 3.5. Our TESs for CNs (2.42±0.24) are comparable to those reported in
41 Knoblauch, Saunders (17). Error score pattern of one CVD participant (CVD#5) is shown in the
42 right panel (note radial axis range is 2-16). Patterns for each CVD observer are provided in Figure
43 S1. TES of 11% CVD observers exceeded the error score range of our CN group while 89% CVDs
44 have TES larger than 100. The right-half MP contains the median of error scores of cap sequence

1 43 to 84, and was used to identify type of defect. The classification criteria are indicated in the
 2 right panel of Figure 1 according to the FM100 manual: protans, deutans, and tritans have right-
 3 half mid-points in ranges 62~70, 56~61, or 46~52, respectively. TES and MP scores of the CVD
 4 group are reported in Table S2. Average time taken for testing participants were 13'32" for 18
 5 CNs and 13'23" for 17 CVDs. As for plotting the pattern and calculating TES and MP manually,
 6 3'17" for CNs and 14'48" for CVDs were taken on average. Computer-based data analysis
 7 algorithms can largely improve efficiency of FM100 test by reducing the data processing time.
 8
 9



10
 11 Figure 1: FM100 hue test results. Left: average error score pattern of 19 CNs. Hues of colored caps
 12 are numbered from 1 to 85 with the corresponding mean error score (black line) and standard error
 13 range indicated along radial coordinates for each hue. Upper and lower standard error ranges are
 14 depicted by the red and blue dashed lines, respectively. The outermost circle where the color dots
 15 reside represents an error score of 3.5, and the center error score is 2, indicating the lowest possible
 16 error score. Right: error score pattern (black line) of an example CVD observer (CVD#5). Note
 17 that the largest radial scale is 16. Mid-point and TES of this observer as well as diagnostic curves
 18 (color arcs) are also shown. Mid-point of this observer falls in the protan range.
 19

20 *Rayleigh Color Match*

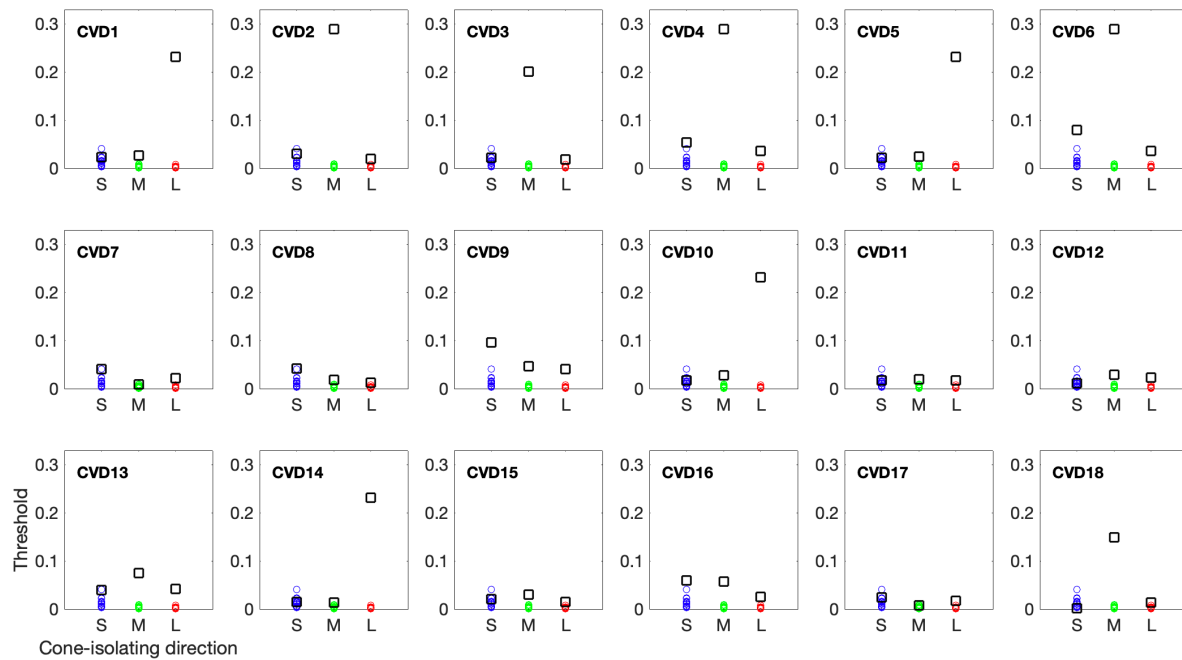
21 Testing time duration for Rayleigh match on an anomaloscope was 10.64 min on average for 14
 22 CNs to complete 8 measurements (4 with each eye) and 15.07 min for 6 CVDs to complete all 8
 23 measurements with an additional matching range identification trial. Detailed testing procedures
 24 and results are reported in the Supplementary material. All 14 CN participants produced normal
 25 anomalous quotients and all 8 CVDs produced atypical quotients. 6 CVD classifications agreed
 26 with HRR classification and 4 agreed with FM100 classifications (Table 1).
 27

28 *FInD Color detection and discrimination thresholds and K-means clustering*

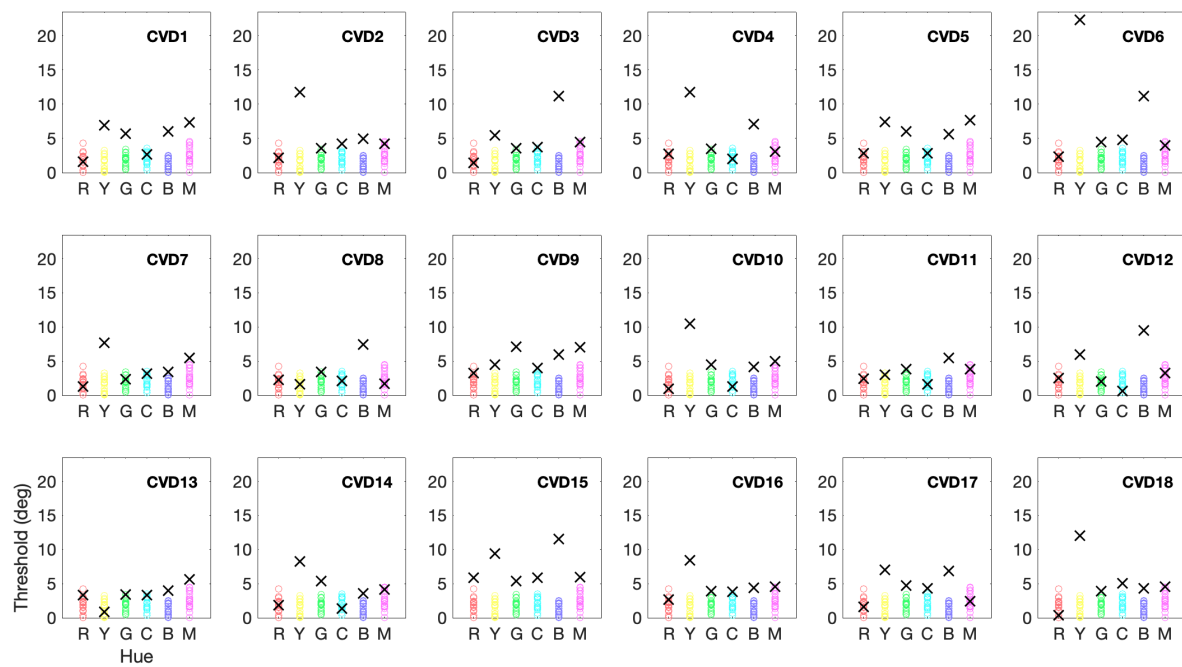
29 FInD Color detection thresholds for each CVD observer are shown in the top panels in Figure 2,
 30 and low-saturation discrimination thresholds are shown in bottom panels (Figure 2b; high-
 31 saturation discrimination thresholds are shown in Figure S2). For both tasks, data of CNs tend to
 32 cluster at lower thresholds and have relatively small variance, whereas there are large individual

1 differences among CVD observers with different patterns and degrees of selective threshold
2 elevation likely corresponding to CVD types. Average (\pm standard deviation) time durations taken
3 for CN and CVD participants to complete all trials in the detection task were 4'57" (\pm 1'59") and
4 5'08" (\pm 1'45"), and were 19'57" (\pm 5'33") and 18'04" (\pm 5'32") to complete all trials in the
5 discrimination task, respectively. Given that two saturation levels were investigated in the
6 discrimination task, one test level takes only half of the time reported (less than 10 min).
7

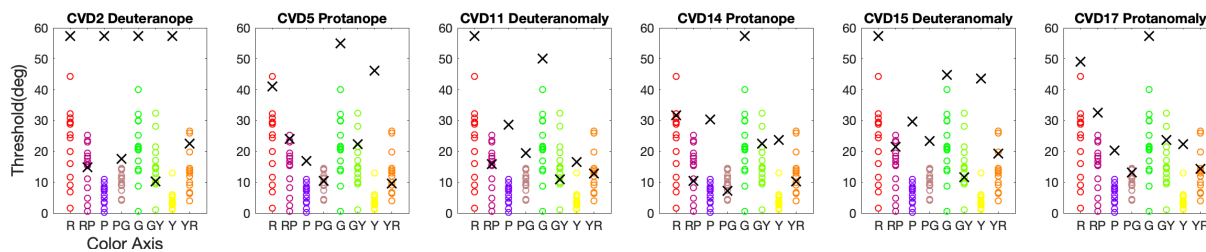
(a) FinD Color detection thresholds



(b) FinD HSV Color discrimination thresholds



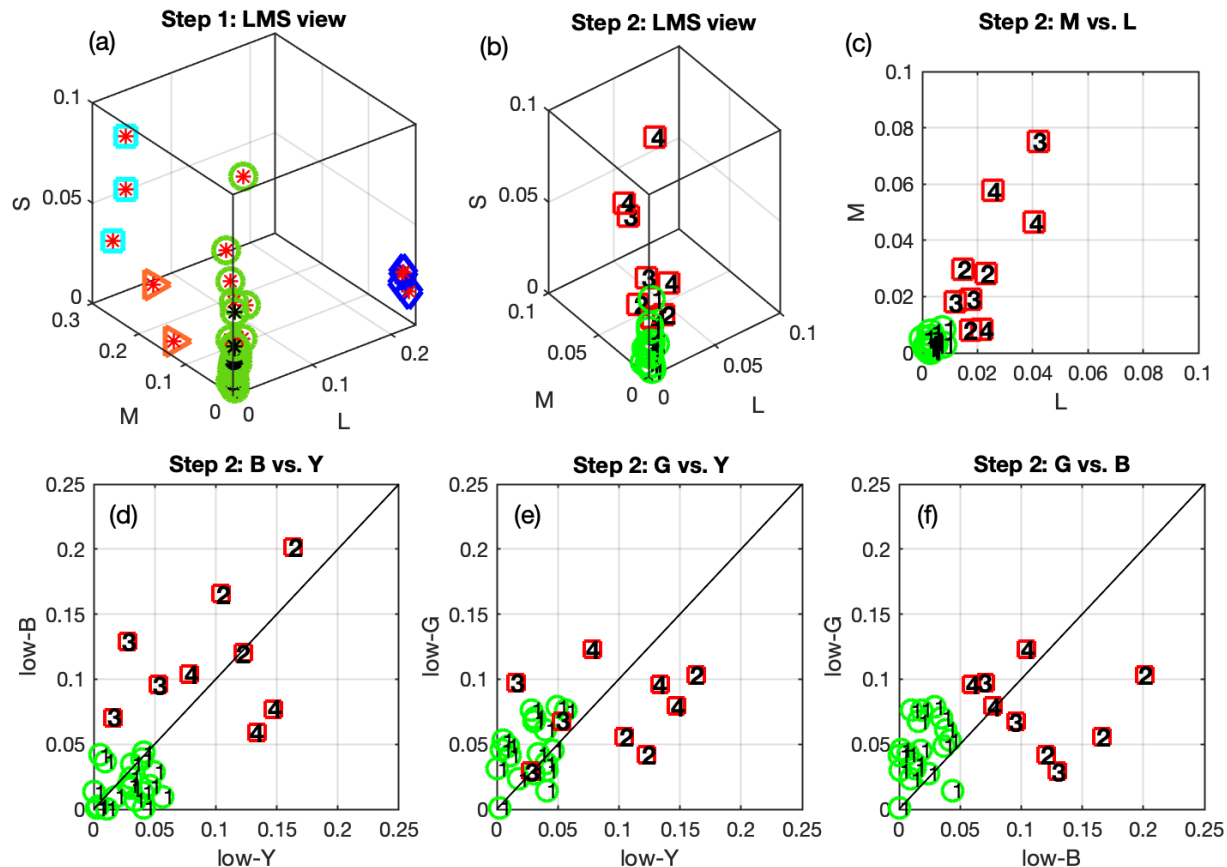
(c) FinD Equiluminant Color discrimination thresholds



1
2 Figure 2: FInD Color detection and discrimination thresholds. Thresholds of all CN participants
3 are plotted as colored circles in all panels as references, and thresholds of each CVD observer are
4 denoted by black squares (detection) or crosses (discrimination) in separate panels. (a) FInD Color
5 detection thresholds (upper panels) are plotted as cone contrast vector length and (b) low-saturation
6 discrimination (lower panels) thresholds are plotted in degrees of HSV color space angle. (c) shows
7 results of 6 CVD participants for the FInD Color discrimination task with equiluminant, colored
8 stimuli.

9
10 To set up diagnostic criteria for FInD Color results, we sought to achieve an automated
11 classification of CVD type and severity using unsupervised machine learning (UML). To that end,
12 a K-means clustering algorithm was applied with inputs being subsets of the detection and
13 discrimination threshold datasets. A two-step classification was performed (see details in
14 Supplementary section: *K-means classification*). The first step took LM detection thresholds as
15 inputs and was able to segregate a large group. Observers in this group include CN and potential
16 anomalous trichromats (AT) with low LM thresholds (Figure 3a green circles). We refer to this
17 group of participants who have defective color vision but low detection thresholds as ‘potential
18 AT’ for simplicity in the following paragraphs. Three smaller clusters were also identified, one
19 with high L and low M thresholds (CVD#1,5,14; likely protanopes; Figure 3a blue diamonds), one
20 with high M and low L thresholds (CVD#2,4,6; likely deuteranopes; Figure 3a cyan squares), and
21 one with intermediate M and low L thresholds (CVD#3,18; likely extreme deuteranomaly or
22 deuteranopes; Figure 3a orange triangles). Observers who were classified in the three small groups
23 received consistent classifications as from HRR, FM100, and anomaloscope data if available
24 (Table 1). The second classification step took LMS detection thresholds and low-saturation yellow
25 (Y), blue (B) and magenta (G) hue discrimination thresholds as inputs, and divided the large group
26 from the first step to four smaller clusters: one CN group (Figure 3b-f: “1”s surrounded by green
27 circles), one deutan group (likely deuteranomaly; Figure 3 b-f: “3”s surrounded by red squares; 2
28 received consistent HRR and FM100 classifications, and 1 received inconsistent HRR and FM100
29 classifications but was classified as deuteranomaly by the anomaloscope), one protan group (likely
30 protanomaly; Figure 3 b-f: “4”s surrounded by red squares; 2 received consistent HRR and FM100
31 classifications, and 1 received inconsistent HRR and FM100 classifications) and one unknown AT
32 group (Figure 3 b-f: “2”s surrounded by red squares; 2 received inconsistent HRR and FM100
33 classifications, and were classified as protanomaly and deuteranomaly, respectively, by the
34 anomaloscope, and 1 received consistent HRR and FM100 classifications as protan). With the lack
35 of genetic measurements available for the participants, however, we are not able to relate the
36 subgroups to any specific genotype.

37
38



1
 2 Figure 3. Classification of CN and CVD results. (a) Step one clustering results illustrated in LMS
 3 detection threshold space. All 37 individual data points are shown. CVD and CN individuals are
 4 represented by red and black asterisks, respectively. Four clusters, denoted in differently shaped
 5 and colored symbols (green circles, blue diamonds, cyan squares, and orange triangles) around the
 6 asterisks, are identified. (b)-(f) show step two clustering results. Only individual points surrounded
 7 by green circles ($n=28$) in (a) are taken and plotted. These thresholds were clustered as four groups
 8 denoted by numbers. Red squares and green circles represent CVD and CN, respectively.

9
 10 *Test-retest reliability*

11 Thirteen CN and eight CVD participants completed the retest session. All CNs passed HRR on
 12 both tests. HRR categorization of two CVD observers were changed: one observer (CVD#14)
 13 stayed the same type (protan) but with a different severity (medium to strong), and the other
 14 observer (CVD#12) changed in both type and severity (strong protan to medium deutan). Bland-
 15 Altman analyses(18) show no significant learning effect and bias (see details in the Supplementary
 16 material).

17
 18 *Comparison of the methods*

19 Classification results with three methods for the 18 CVD observers are compared in Table 1. As
 20 we used consistent Rayleigh match, HRR and FM100 CVD type classification as references in the
 21 two-step unsupervised classification, agreement between FInD Color and the other tasks should
 22 be expected. For the observers classified as strong or medium protan or deutan (likely protanope
 23 or deutanope) using FInD Color LM detection thresholds, CVD type classification of the three

1 methods agree perfectly (CVD#1,2,3,4,5,6,14,18). However, LM detection thresholds alone were
 2 not able to distinguish between CN and AT as they both have relatively low cone specific detection
 3 thresholds. With the addition of the hue discrimination thresholds, the algorithm successfully
 4 distinguished CN observers from anomalous trichromats (AT), and further assign potential AT to
 5 CVD subtypes. The second-step classification results agree for all 19 CN observers and 5 out of 9
 6 AT observers. The remaining four AT observers (CVD#7,11,15,17) received inconsistent
 7 diagnoses from HRR and FM100, therefore for FInD classification, we assign them to the CVD
 8 categories in which the members in their clusters fall, which necessarily agree with one of their
 9 HRR or FM100 diagnoses, but not both. It's worth noting that in one cluster (CVD#:12,15,17),
 10 although one of the participants (CVD#12) received consistent HRR and FM100 diagnoses in the
 11 test session, the retest HRR diagnosis changed type, so all three participants received confusing
 12 HRR and FM100 diagnosis when the retest session results are considered. This cluster is then
 13 referred to as the unknown AT group. As for testing durations, a FInD Color detection task with
 14 three testing directions and three trials per direction (CN: 4'57", CVD: 5'08") took longer than
 15 HRR (CN: 37", CVD: 2'53"), but quick screening with only one trial took shorter (1'43") for CVD.
 16 A single level FInD Color discrimination task (CN: 9'59", CVD: 9'02") was more rapid than
 17 FM100 (CN: 13'32", CVD: 13'23"). FM100 testing can be further prolonged due to the additional
 18 time spent on data processing.

19

CVD#	Anomaloscope	HRR		FM100	FInD Color K-means	
		Type	Severity	Type	Type	Severity
1		Protan	Strong	Protan	Protan	Strong
2	Deuteranopia	Deutan	Strong	Deutan	Deutan	Strong
3		Deutan	Strong	Deutan	Deutan	Medium
4		Deutan	Strong	Deutan	Deutan	Strong
5	Protanopia	Protan	Strong	Protan	Protan	Strong
6		Deutan	Strong	Deutan	Deutan	Strong
7		Pro/Deu	Mild	Deutan	Protan	Mild
8		Deutan	Mild	Deutan	Deutan	Mild
9		Protan	Strong	Protan	Protan	Mild
10		Protan	Strong	Protan	Protan	Strong
11	Deuteranomaly	Deutan	Mild	Protan	Deutan	Mild
12		Protan	Strong	Protan	unknown AT	Mild
13		Deutan	Strong	Deutan	Deutan	Mild
14	Protanopia	Protan	Medium	Protan	Protan	Strong
15	Deuteranomaly	Deutan	Mild	Protan	unknown AT	Mild
16		Protan	Medium	Protan	Protan	Mild
17	Protanomaly	Deutan	Mild	Protan	unknown AT	Mild
18		Deutan	Strong	Deutan	Deutan	Medium

20

21 Table 1. Identification of CVD type and severity for 18 CVD observers for the three methods.

22

23 Discussion

24 The current study introduced and measured performance of the newly designed FInD Color
 25 detection and discrimination tasks, and compared them against HRR, FM100 hue and Rayleigh
 26 match tests. FInD is rapid, self-administered, and easy to use, without strict operating regulations
 27 other than the normal usage of a computer, so can be readily grasped by patients or children(19).

1 The combination of color detection and color discrimination performance also provides richer
2 information about CVD color perception, and enables classification of color vision subtypes using
3 unsupervised machine learning (UML) classification. In comparison, HRR, although rapid and
4 easy, provides limited information about atypical color perception, and is not able to classify CVD
5 type for very mild cases (CVD#7). FM100 measures color discrimination but does not have strict
6 and precise diagnostic metrics. Even in such a small sample, some CVD observers obtained lower
7 error scores than observers with normal color vision but poor discrimination. Moreover, the actual
8 error score pattern shows large degree of variation instead of a clearly classifiable pattern. FInD
9 Color tasks afford a remedy for these shortcomings, which may be critical for detecting and
10 tracking progression or remediation of acquired CVD in neuro-ophthalmic disease(2).
11 Furthermore, FInD discrimination task significantly reduces testing and analysis duration
12 compared to FM100, and FInD detection task, although taking similar amount of time compared
13 to HRR for CVD testing, is more informative.

14
15 In the present implementation of FInD Color detection, we calibrated the monitor to generate and
16 test detection thresholds for short-, medium- and long-wavelength sensitive cone-isolating stimuli,
17 based on standard cone fundamentals (see Methods). For FInD Color discrimination, we employed
18 HSV color space to generate stimulus pairs of differing colors. HSV is based on subjective
19 estimates of distance in perceptual color space, which is convenient for measurement of color
20 perception on standard computer displays. However, different color hues in HSV may not be
21 equiluminant, and therefore participants could utilize luminance differences instead of color
22 differences to select cells that contained blobs of differing color. We attempted to mask any such
23 cues with 20% contrast dynamic luminance noise, and in the Supplementary materials we report
24 the magnitude of luminance artifacts in our display. The results show that artifacts differ with test
25 hue angle in a pattern that is inconsistent with differences in discrimination threshold and are less
26 than 6 cd/m^2 for all test axes at the highest discrimination difference for CN participants. A
27 separate luminance-matched achromatic blob discrimination task was conducted to evaluate
28 magnitude of luminance artifacts, and thresholds for luminance difference are reported in Table
29 S7. Sensitivities of CN observers to low-saturation luminance differences never exceed the artifact
30 generated at a threshold level of 5° hue angle. However, Sensitivities to high-saturation luminance
31 differences along three hue axes (R, C, G) exceed the threshold at 5° hue angle. In this case, the
32 results of the FInD Color discrimination test are conservative, and we would expect color
33 discrimination thresholds to be even higher for CVD participants. To assess performance of the
34 discrimination task in a color space free of luminance artefacts, a subset of participants (6 CVDs
35 and 14 CNs) were available to complete the discrimination task but with colors chosen from an
36 equiluminant plane (See Figure 2(c), and details in Supplementary, *FInD Color discrimination*
37 *task with an equiluminant color plane*). Eight color axes were measured. Results show that
38 response patterns using the equiluminant plane are similar to that of the HSV space: both protans
39 and deutans have significantly elevated thresholds at yellow, green, blue/purple axes, with
40 thresholds for red and other intermediate axes being more variable (see p-value tables in
41 Supplementary).

42
43 Widely used CVD categories adopted by many color vision screening tests were predicated on
44 color vision phenotype which allows rapid detection of abnormal performance. Albeit simple, the
45 conventional categories disguise the variable nature of CVD. The well-known large individual
46 variations in CVD make the diagnostic groups less distinguishable, and their relationship to

1 genotypes remains unclear(20, 21). For instance, in addition to the protanomaly category, one
2 CVD type (Pseudo-protanomaly), while having LM photopigment peak similar to protanopes, is
3 able to produce trichromacy with differed optical density(22). The story can become more
4 complicated when interactions between postreceptoral signals are considered. Brain plasticity
5 seems to elicit compensatory adjustments so that the actual perceptual color ability loss for CVD
6 can be reduced, further blurring boundaries between discrete phenotypes(23). UML has the ability
7 to group individuals with similar features to the same category, and has previously been used to
8 successfully identify subtypes of other diseases(24). Finding categories for groups is a non-trivial
9 challenge for the following reasons: using prior information such as self-report of CVD or clinical
10 results to classify data is dependent on the methods used in the clinical tests; Furthermore, treating
11 CN and CVDs as one distribution and using multiple standard deviations away from the mean of
12 CNs does not consider the known CVD subtypes that belong in separate distributions. Hence, we
13 decided to use UML technique — k-means to address these problems to classify groups without
14 assigning a group *a priori*. The novel application of using k-means clustering algorithm to establish
15 CVD classification criteria in the present work, sheds light on the competence of machine learning
16 techniques in capturing obscure response patterns in psychophysical measurements of typical and
17 atypical color vision, thus possessing the potential to reveal the continuous and variable nature of
18 CVD. As a result, the CVD group can be better categorized into subtypes, and inconclusive CVD
19 cases that received confusing diagnoses from traditional tests can be better understood and treated.

20
21 The addition of hue discrimination thresholds in the second-step classification was essential to
22 discern potential ATs, as detection thresholds alone did not effectively differentiate potential
23 anomalous trichromats from normal trichromats (Note that the LMS cone isolating directions used
24 in the detection task are computed based on a standard observer with normal color vision, therefore
25 certain degree of individual differences should always be considered). This finding suggests
26 complexity in AT phenotypes, echoing the previous evidence(20). We caution that inferences
27 about AT features made based upon current results should remain provisional until larger datasets
28 are acquired, which would also benefit the comprehensive evaluation of the FInD method. As
29 discussed above, even though the HSV plane in the discrimination task is not equiluminant, our
30 classification results demonstrate its ability to assay response pattern differences between CN and
31 CVD, with significantly different discrimination patterns observed and the pattern of findings was
32 the same for an equiluminant color space. Apart from HSV, the FInD tasks implemented in this
33 study required careful display calibration to generate cone-isolating directions (for an ideal
34 observer model) and equiluminant color planes. The same concerns prevail for the anomaloscope
35 and for the light source and pigment decay of HRR and FM100 tests. Therefore, deployment of
36 these approaches will depend on calibration of the use of specific model specifications. The ability
37 of FInD Color tasks to detect and classify tritan deficiencies, inherited or acquired, was not tested
38 in the current study. Future efforts are needed to verify whether the unused features, S detection
39 and the rest of the hue discrimination thresholds, are critical for the categorization of tritans.

40
41 In summary, this proof-of-concept study has shown that FInD Color tasks can provide continuous
42 color detection and discrimination threshold estimates that may track change in CVD and could
43 therefore serve as a rapid and easy-to-use tool for clinical monitoring and diagnosis. The HSV
44 colors are equally diagnostic compared to the equiluminant colors. FInD also has potential in basic
45 color vision research, in that the combination of FInD Color tasks with UML technique might

1 provide insights about hidden structures in the data and further assist the understanding of
2 defective color mechanisms.

5 **Acknowledgements**

6 The work is supported by NIH Grant R01 EY029713.

8 **Competing interests**

9 FInD is patented & owned by Northeastern University, USA. JS & PJB are founders of PerZeption
10 Inc., to which the FInD method is exclusively licensed. JH declares that no competing interests
11 exist.

14 **References**

- 15 1. Sharpe LT, Stockman A, Jägle H, Nathans J. Opsin genes, cone photopigments, color
16 vision, and color blindness. *Color vision: From genes to perception*. 1999;351:3-52.
- 17 2. Simunovic MP. Acquired color vision deficiency. *Survey of Ophthalmology*.
18 2016;61(2):132-55.
- 19 3. Zabel J, Przekoracka-Krawczyk A, Olszewski J, Michalak KP. Variability of Rayleigh
20 and Moreland test results using anomaloscope in young adults without color vision disorders.
21 *Plos one*. 2021;16(5):1-15.
- 22 4. Bailey JE, Neitz M, Tait DM, Neitz J. Evaluation of an updated HRR color vision test.
23 *Visual Neuroscience*. 2004;21(3):431-6.
- 24 5. Cole BL, Lian Ky, Lakkis C. The new Richmond HRR pseudoisochromatic test for
25 colour vision is better than the Ishihara test. *Clinical and Experimental Optometry*.
26 2006;89(2):73-80.
- 27 6. Evans BE, Rodriguez-Carmona M, Barbur JL. Color vision assessment-1: Visual signals
28 that affect the results of the Farnsworth D-15 test. *Color Research & Application*. 2021;46(1):7-
29 20.
- 30 7. Mollon JD, Regan BC. *Cambridge Colour Test Handbook*. (Cambridge Research
31 Systems Ltd, 2000).
- 32 8. Barbur JL, Harlow J, Plant GT. Insights into the different exploits of colour in the visual
33 cortex. *Proceedings of the Royal Society of London Series B: Biological Sciences*.
34 1994;258(1353):327-34.
- 35 9. Tang T, Alvaro L, Alvarez J, Maule J, Skelton A, Franklin A, et al. ColourSpot, a novel
36 gamified tablet-based test for accurate diagnosis of color vision deficiency in young children.
37 *Behav Res Methods*. 2022;54(3):1148-60.
- 38 10. Bex P, Skerswetat J. FInD-Foraging Interactive D-prime, a rapid and easy general
39 method for visual function measurement. *Journal of Vision*. 2021;21(9):2817.
- 40 11. Merabet LB, Manley CE, Pamir Z, Bauer CM, Skerswetat J, Bex PJ. Assessing Motion
41 and Form Coherence Processing in Cerebral Visual Impairment. *Developmental Medicine &*
42 *Child Neurology*. In Press.
- 43 12. Rabin J, Gooch J, Ivan D. Rapid quantification of color vision: the cone contrast test.
44 *Investigative Ophthalmology & Visual Science*. 2011;52(2):816-20.
- 45 13. Kleiner M, Brainard DH, Pelli D. What's new in Psychtoolbox-3? *Perception*.
46 2007;36(Suppl.):14.

- 1 14. Stockman A, Sharpe LT. The spectral sensitivities of the middle- and long-wavelength-
2 sensitive cones derived from measurements in observers of known genotype. *Vision Research*.
3 2000;40(13):1711-37.
- 4 15. Stockman A, Sharpe LT, Fach C. The spectral sensitivity of the human short-wavelength
5 sensitive cones derived from thresholds and color matches. *Vision research*. 1999;39(17):2901-
6 27.
- 7 16. He J, Taveras-Cruz Y, Eskew RT, Jr. Modeling individual variations in equiluminance
8 settings. *Journal of Vision*. 2021;21(7):1-16.
- 9 17. Knoblauch K, Saunders F, Kusuda M, Hynes R, Podgor M, Higgins KE, et al. Age and
10 illuminance effects in the Farnsworth-Munsell 100-hue test. *Applied Optics*. 1987;26(8):1441-8.
- 11 18. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat*
12 *Methods Med Res*. 1999;8(2):135-60.
- 13 19. Merabet LB, Manley CE, Pamir Z, Bauer CM, Skerswetat J, Bex PJ. Motion and form
14 coherence processing in individuals with cerebral visual impairment. *Dev Med Child Neurol*.
15 2023;00:1-8.
- 16 20. Bosten J. The known unknowns of anomalous trichromacy. *Current Opinion in*
17 *Behavioral Sciences*. 2019;30:228-37.
- 18 21. Neitz J, Neitz M. The genetics of normal and defective color vision. *Vision Res*.
19 2011;51(7):633-51.
- 20 22. Neitz J, Neitz M, He JC, Shevell SK. Trichromatic color vision with only two spectrally
21 distinct photopigments. *Nature Neuroscience*. 1999;2(10):884-8.
- 22 23. Isherwood ZJ, Joyce DS, Parthasarathy MK, Webster MA. Plasticity in perception:
23 insights from color vision deficiencies. *Faculty Reviews*. 2020;9:8.
- 24 24. Eshaghi A, Young AL, Wijeratne PA, Prados F, Arnold DL, Narayanan S, et al.
25 Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data.
26 *Nature communications*. 2021;12(1):1-12.
27