



OPEN

## Quantification of expected information gain in visual acuity and contrast sensitivity tests

Zhong-Lin Lu<sup>1,2,3✉</sup>, Yukai Zhao<sup>4</sup>, Luis Andres Lesmes<sup>5</sup> & Michael Dorr<sup>5</sup>

We make use of expected information gain to quantify the amount of knowledge obtained from measurements in a population. In the first application, we compared the expected information gain in the Snellen, ETDRS, and qVA visual acuity (VA) tests, as well as in the Pelli–Robson, CSV-1000, and qCSF contrast sensitivity (CS) tests. For the VA tests, ETDRS generated more expected information gain than Snellen. Additionally, the qVA test with 15 rows (or 45 optotypes) generated more expected information gain than ETDRS, whether scored with VA threshold alone or with both VA threshold and VA range. Regarding the CS tests, CSV-1000 generated more expected information gain than Pelli–Robson, and the qCSF test with 25 trials generated more expected information gain than CSV-1000, whether scored with AULCSF or with CSF at six spatial frequencies. The active learning-based qVA and qCSF tests have the potential to generate more expected information gain than traditional paper chart tests. Although we have specifically applied it to compare VA and CS tests, expected information gain is a general concept that can be used to compare measurements in any domain.

Measurement serves as the bedrock of the scientific method and finds application across a diverse range of disciplines, encompassing physics, biology, engineering, social science, and medicine<sup>1–4</sup>. It provides the empirical data essential for hypotheses testing, theory formulation, performance evaluation, and diagnostic procedures<sup>5–9</sup>. Despite the development of numerous quality metrics designed to assess individual measurements and the concordance between two unidimensional measurements<sup>10–12</sup>, the task of comparing innovative measurements to the established gold standards remains a formidable challenge. This challenge is particularly pronounced when the novel measurements yield outcomes of distinct or greater dimensionality. Furthermore, conventional metrics, such as those gauging test–retest variability, primarily quantify the uncertainties associated with measurement outcomes and do not encompass an evaluation of the knowledge accrued through the act of measurement.

Viewed through the lens of information theory<sup>13–15</sup>, measurement is conceptually defined as “a set of observations that reduce uncertainty, with the result expressed as a quantity”<sup>16</sup>. In order to precisely quantify the extent of new knowledge derived from measurement, we advocate for the application of expected information gain, also referred to as expected mutual information<sup>17</sup>. While the concept of information gain<sup>18</sup> has found extensive use in machine learning, where it serves as a fundamental criterion for optimizing decision trees<sup>19</sup> and guides the selection of optimal stimuli in active learning algorithms<sup>20–22</sup>, its applications have largely centered on assessing relative information gain among different features or stimuli in the learning process. However, in this particular investigation, we employ it to quantitatively assess the amount of information acquired through measurements conducted on a population.

### Computing the expected information gain

Two equivalent methodologies exist for calculating the expected information gain. The first approach hinges on the reduction of uncertainty regarding the property being measured, often referred to as the “truth”, following the measurement. In contrast, the second approach is grounded in the disparity between the uncertainty associated with all conceivable measurement outcomes within a population and the projected remaining uncertainty after the execution of a measurement. While the first approach is a direct derivation from the concept of expected information gain, the second approach, in practice, tends to be more straightforward to implement. Both approaches are presented in the following section, and additional verification of their equivalence can be located in Appendix 1.

<sup>1</sup>Division of Arts and Sciences, NYU Shanghai, Shanghai, China. <sup>2</sup>Center for Neural Science and Department of Psychology, New York University, 4 Washington Place, New York, NY 10003, USA. <sup>3</sup>NYU-ECNU Institute of Brain and Cognitive Neuroscience at NYU Shanghai, Shanghai, China. <sup>4</sup>Center for Neural Science, New York University, New York, USA. <sup>5</sup>Adaptive Sensory Technology Inc., San Diego, CA, USA. ✉email: zhonglin@nyu.edu

The first approach to calculate expected information gain begins by considering the property to be measured, referred to as the “truth”. This property is represented by a random variable  $X$ , with a probability density function denoted as  $P(x)$  (Fig. 1a). We also consider all possible outcomes of the measurement, represented by a random variable  $Y$ , with its own probability density function  $P(y)$  (Fig. 1c). Following each measurement, the outcome is represented by a probability distribution  $P(y|x)$  (Fig. 1b). By applying Bayes’ rule, we can derive the posterior distribution of  $x$ , denoted as  $P(x|y)$ (Fig. 1d), using the formula:  $P(x|y) = P(y|x)P(x)/P(y)$ , where  $P(y) = \int_X P(y|x)P(x)dx$ . Shannon entropy is then employed to quantify the level of uncertainty associated with  $X$  before any measurement:

$$H(X) = - \int_X P(x) \log_2(P(x)) dx. \tag{1}$$

Following a measurement, the expected entropy of  $X$  is determined as:

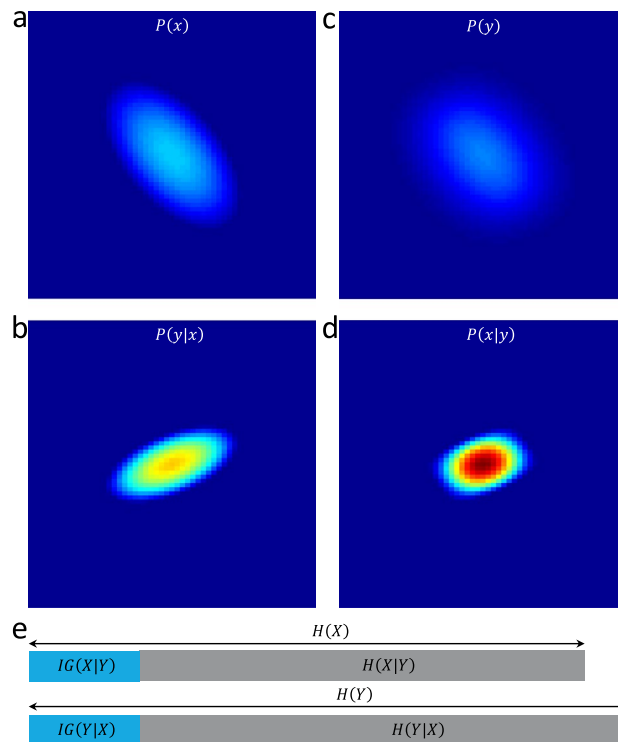
$$H(X|Y) = - \int_Y \left( \int_X P(x|y) \log_2(P(x|y)) dx \right) P(y) dy, \tag{2}$$

and subsequently, the expected information gain is computed (Fig. 1e):

$$IG(X|Y) = H(X) - H(X|Y) = - \int_X P(x) \log_2(P(x)) dx + \int_Y \left( \int_X P(x|y) \log_2(P(x|y)) dx \right) P(y) dy. \tag{3}$$

This approach quantifies the reduction in uncertainty regarding the property  $X$  as a result of the measurement outcomes represented by  $Y$ .

The second approach to compute expected information gain also begins by considering probability density functions:  $P(x)$ ,  $P(y)$  and  $P(y|x)$  (Fig. 1a–c). Using Shannon entropy, we first assess the level of uncertainty associated with  $Y$  based on the distribution of all possible measurement outcomes:



**Figure 1.** (a) Probability distribution of a quantity  $x$  occurring in population  $X$ ,  $P(x)$ . (b) Probability distribution of obtaining measurement  $y$  given  $x$  is the true value being measured:  $P(y|x)$ . (c) Probability distribution of obtaining a measurement  $y$  from all potential measurement outcomes  $Y$  regardless of the underlying true value,  $P(y)$ . (d) Posterior distribution of  $X$ ,  $P(x|y)$ , following a measurement outcome  $y$ . (e) Expected information gain  $IG(X|Y)$  is the difference between  $H(X)$  and the expected posterior entropy  $H(X|Y)$ . Expected information gain  $IG(Y|X)$  is the difference between  $H(Y)$  and the expected residual entropy  $H(Y|X)$ .  $IG(X|Y) = IG(Y|X)$ .

$$H(Y) = - \int_Y P(y) \log_2(P(y)) dy. \tag{4}$$

Next, we determine the expected residual uncertainty of Y after a measurement:

$$H(Y|X) = - \int_X \left( \int_Y P(y|x) \log_2(P(y|x)) dy \right) P(x) dx. \tag{5}$$

Finally, the expected information gain (Fig. 1e) is computed as:

$$IG(Y|X) = H(Y) - H(Y|X) = - \int_Y P(y) \log_2(P(y)) dy + \int_X \left( \int_Y P(y|x) \log_2(P(y|x)) dy \right) P(x) dx. \tag{6}$$

This second approach assesses how much uncertainty remains in Y after obtaining the measurement outcomes represented by  $P(y|x)$ . In essence, it quantifies the reduction in uncertainty associated with Y as a result of measuring X. Both approaches yield equivalent results and provide valuable insights into the information gained through measurements in different ways.

### A practical illustration of expected information gain

We present a practical illustration of expected information gain using two rulers, considering both the first (Fig. 2a,d) and second (Fig. 2b,c) approaches:

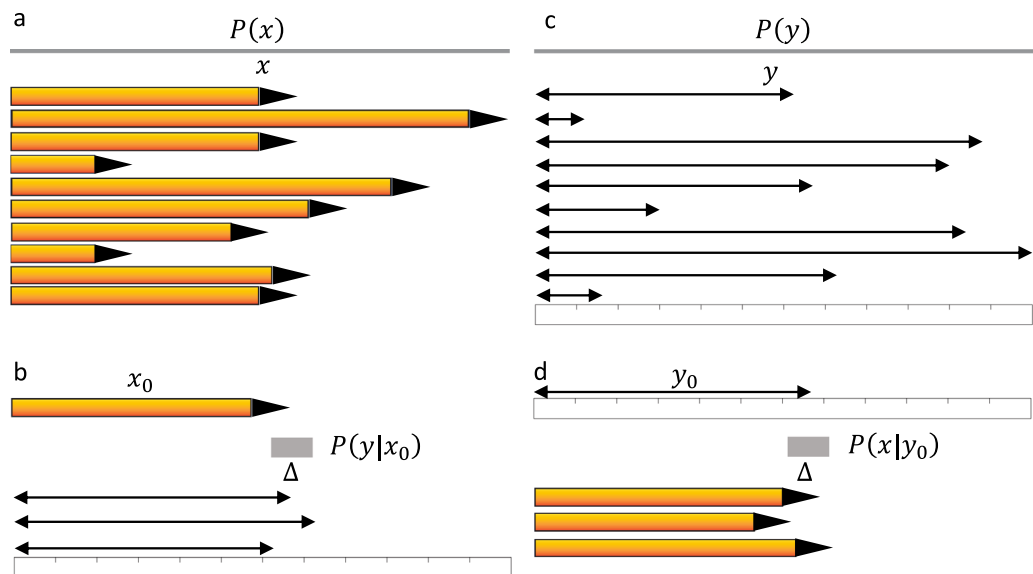
- *First approach* For a ruler with a unit  $\Delta$ , the probability distribution of measuring an object with length  $x$  is represented by  $P(y|x) = U(x - \Delta/2, x + \Delta/2)$ , where  $U(m, n)$  is a uniform distribution with boundaries  $m$  and  $n$ . If this ruler is used to measure objects with lengths between 0 and  $L$  with equal probability, the probability distribution for object lengths is  $P(x) = U(0, L)$ . The entropy of X before any measurement  $H(X)$  is:

$$H(X) = - \int_0^L U(0, L) \log_2(U(0, L)) dx = \log_2(L). \tag{7a}$$

The outcome distribution  $P(y)$  can also be determined:

$$P(y) = \int_0^L U\left(x - \frac{\Delta}{2}, x + \frac{\Delta}{2}\right) U(0, L) dx = U(0, L). \tag{7b}$$

The entropy of X after considering Y is:



**Figure 2.** An illustration of expected information gain from measurement of the length of pencils using a ruler with unit  $\Delta$ . (a) A uniformly distributed pencil length X,  $P(x)$ . (b) Outcome distribution from measuring the length of a pencil with true length  $x_0$ ,  $P(y|x_0)$ . (c) A uniformly distributed outcomes Y,  $P(y)$ . (d) Posterior distribution of  $x$  based on measurement outcome  $y_0$ ,  $P(x|y_0)$ .

$$H(X|Y) = - \int_0^L \int_0^L U\left(x - \frac{\Delta}{2}, x + \frac{\Delta}{2}\right) \log_2\left(U\left(x - \frac{\Delta}{2}, x + \frac{\Delta}{2}\right)\right) dx U(0, L) dy = \log_2(\Delta). \quad (7c)$$

The expected information gain is then computed as:

$$IG(X|Y) = H(X) - H(X|Y) = \log_2(L/\Delta). \quad (7d)$$

- *Second approach* The entropy of Y ( $H(Y)$ ) is calculated:

$$H(Y) = - \int_0^L U(0, L) \log_2(U(0, L)) dy = \log_2(L). \quad (7e)$$

The expected residual uncertainty of Y after considering X is determined:

$$H(Y|X) = - \int_0^L \int_0^L U\left(x - \frac{\Delta}{2}, x + \frac{\Delta}{2}\right) \log_2\left(U\left(x - \frac{\Delta}{2}, x + \frac{\Delta}{2}\right)\right) dy U(0, L) dx = \log_2(\Delta). \quad (7f)$$

The expected information gain is then computed as:

$$IG(Y|X) = H(Y) - H(Y|X) = \log_2(L/\Delta). \quad (7g)$$

Consider two one-foot-long rulers, one with a one-inch unit and the other with a 1/16-inch unit. Using both approaches, the expected information gains are calculated as  $\log_2(12/1) = 3.58$  bits and  $\log_2(12/(1/16)) = 7.58$  bits for the two rulers, respectively. These values correspond to 12 and 192 distinct length classes, consistent with the number of units on the rulers. These calculations demonstrate how expected information gain can be applied to assess the knowledge gained from measurements using different rulers with varying units of measurement. It provides a quantitative measure of the information acquired through these measurements, which can be valuable in various contexts.

### Expected information gain for a unidimensional measurement with a normal outcome distribution

We next show how to derive the expected information gain for a unidimensional measurement with a normal outcome distribution and provide an intuitive interpretation. In this case, the probability distribution for measuring  $y$  given  $x$  is defined as  $P(y|x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu(x))^2}{2\sigma^2}}$ , where  $\mu(x)$  is the expected value of  $x$ . If both  $x$  and  $y$  are uniformly distributed in an interval of length  $L$ , then  $H(X) = \log_2(L)$ ,  $H(X|Y) = \log_2(\sigma\sqrt{2\pi}e)$ ,  $H(Y) = \log_2(L)$ ,  $H(Y|X) = \log_2(\sigma\sqrt{2\pi}e)$ , and  $IG(X|Y) = IG(Y|X) = \log_2\left(\frac{L}{4.13\sigma}\right)$ . Intuitively, these values indicate the following: (1)  $L/(4.13\sigma)$  classes: This represents the number of distinct classes or intervals into which the measurement outcomes can be categorized. Each class has a size of approximately  $4.13\sigma$ , which corresponds to the 98% confidence interval of the outcome distribution. In other words, this is a way to quantitatively express the granularity or resolution of the measurement. (2) The 98% Confidence Interval: The size of each class,  $4.13\sigma$ , represents the range within which an observed measurement is likely to fall with a high level of confidence (98% confidence interval). This interval provides a measure of uncertainty associated with the measurement outcome.

**Overview of the current study**

Expected information gain, which quantifies the knowledge gained through measurement, is not limited to unidimensional examples. It can be effectively used to compare measurements with any level of dimensionality. In this study, we extended its application to the assessment of visual acuity (VA) and contrast sensitivity (CS) tests, which inherently involve measurements with different dimensionalities. By applying expected information gain to VA and CS tests, we aimed to provide a quantitative basis for comparing these tests, considering their varying optotypes, outcome dimensionalities, and their ability to provide valuable knowledge.

### Overview of the current study

Visual acuity is a crucial measure of visual function and is widely used for diagnosing and managing visual diseases, evaluating the effectiveness of treatments, and establishing professional standards<sup>23–25</sup>. The gold standard test, the ETDRS chart<sup>26</sup>, consisting of rows of five equal-sized optotypes, is used to generate a unidimensional VA threshold score for each patient. However, a newer test called the qVA test<sup>27,28</sup> is introduced, which involves three equal-sized optotypes in each trial and generates a two-dimensional score, incorporating VA threshold and VA range. Importantly, this added dimension provides a more comprehensive assessment of visual acuity, and not considering it could lead to an incomplete evaluation<sup>28</sup>.

The contrast sensitivity function is an increasingly important measure in clinical research and clinical trials as it offers a more comprehensive characterization of spatial vision compared to VA<sup>29,30</sup>. Multiple instruments have been developed to measure CS<sup>20,31,32</sup>, each using different optotypes and producing scores with varying dimensionalities. For example, the Pelli–Robson test<sup>32</sup> uses unfiltered Sloan letter stimuli and generates a unidimensional CS score at one spatial frequency, while the CSV-1000 test<sup>33</sup> uses windowed sinewave grating stimuli and provides a four-dimensional CS score at four spatial frequencies. Another test, the qCSF<sup>20,34</sup>, uses filtered Sloan letter stimuli and generates both a unidimensional area under the log CSF (AULCSF) score and

a six-dimensional CS score at six spatial frequencies. This diversity in CS tests and their outcome dimensions poses challenges when comparing their effectiveness.

While many studies have assessed the accuracy, test–retest variability, sensitivity, and specificity of VA and CS tests<sup>28,35–47</sup>, comparing them has proven difficult due to differences in optotypes, optotype arrangements, and outcome dimensionalities. Existing metrics often focus solely on the uncertainty of measurement outcomes and do not quantify the knowledge gained through these measurements. This study aimed to address these challenges by employing the concept of expected information gain to compare VA and CS tests. Computer simulations were used to calculate and compare the expected information gain and the number of classes derived from the Snellen chart<sup>48</sup>, ETDRS chart<sup>26</sup>, and qVA test<sup>27</sup> for VA assessment, as well as the Pelli–Robson chart<sup>32</sup>, CSV-1000 chart (Vector Vision, Houston, Texas)<sup>33</sup>, and qCSF test<sup>20,34</sup> in populations with uniform distributions, as well as populations with distributions based on experimental data. This approach provides a quantitative and knowledge-based method for comparing these essential vision tests.

## Methods

### Apparatus

All the simulations and analyses were conducted on a Dell computer with Intel Xeon W-2145 @ 3.70 GHz CPU (8 cores and 16 threads) and 64 GB installed memory (RAM) with Matlab R2019a (MathWorks Corp., Natick, MA, USA) and R (R Core Team, 2020).

### Visual acuity tests

#### Simulated observers

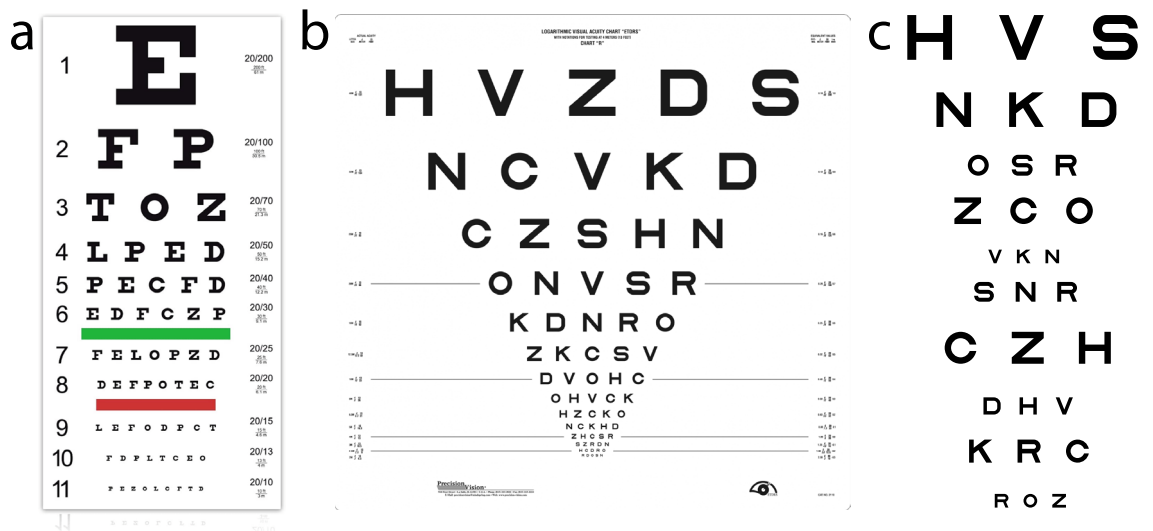
We conducted two simulations using the Snellen, ETDRS, and qVA tests (Fig. 3). In Simulation 1, we simulated 1386 observers from a uniform distribution of VA threshold ( $\theta_{Threshold}^{VA}$ ) and VA range ( $\theta_{Range}^{VA}$ ), with  $\theta_{Threshold}^{VA} \in [-0.3, 1.0] \log_{10} \text{MAR}$  and sampled every 0.02 logMAR, and  $\log_{10}(\theta_{Range}^{VA}) \in [-1.0, 0]$  and sampled every 0.05 logMAR units. In Simulation 2, we simulated 1386 observers from the population distribution of VA threshold and VA range derived from an existing qVA dataset of 14 eyes tested with Bangerter foils<sup>28,49</sup>. The original experiment was conducted at the Ohio State University. Written consent was obtained from all the participants before the experiment. The study protocol was approved by the institutional review board of human subject research of The Ohio State University and adhered to the tenets of the Declaration of Helsinki.

#### Visual acuity behavioral function

For each simulated observer, the discriminability ( $d'$ ) for an optotype of size  $s$  is described by the VA behavioral function (Patent No. US 10758120B2)<sup>27</sup>:

$$\log_{10}(d'(s|\theta^{VA})) = \log_{10}(6) + \frac{\omega}{2\theta_{Range}^{VA}}(s - \theta_{Threshold}^{VA}) - \frac{1}{2} \log_{10} \left( 8 + 10^{\frac{\omega}{\theta_{Range}^{VA}}(s - \theta_{Threshold}^{VA})} \right), \quad (8)$$

where  $\theta^{VA} = (\theta_{Threshold}^{VA}, \theta_{Range}^{VA})$ ,  $\theta_{Threshold}^{VA}$  is VA threshold, corresponding to the optotype size at  $d' = 2$ ,  $\theta_{Range}^{VA}$  is the VA range of the behavioral function, that is, the range of optotype sizes that covers  $d' = 1$  to  $d' = 4$  performance levels, and  $\omega = \log_{10} 35 - \log_{10} 1.25$ .



**Figure 3.** (a) A Snellen chart (Image courtesy of Precision Vision, Inc). (b) An ETDRS chart (Image courtesy of Precision Vision, Inc). (c) A subset of potential stimuli in qVA.

In these VA tests, observers identify optotypes from the 10 Sloan letters. We can compute the probability of obtaining  $m$  correct responses in the 10-alternative forced identification task in a row of  $M$  randomly sampled optotypes from their  $d'$  function (Appendix 2):

$$p(m|s, M, \theta^{VA}) = f\left(m, M, d'(s|\theta^{VA})\right), \quad (9)$$

where  $f(\cdot)$  is derived from signal detection theory by considering chart design, i.e., the number of optotypes in a row and whether optotypes in each row are sampled from the 10 Sloan letters with or without replacement.

#### Snellen chart

The Snellen chart (Fig. 3a) has 11-rows, with 1, 2, 3, 4, 5, 6, 7, 8, 8, and 9 optotypes in each row and optotype size descending from 1.0 to  $-0.3 \log\text{MAR}$ . Each simulated observer was tested with the standard procedure<sup>50</sup>. The probability of correctly identifying  $m$  optotypes in a row is determined by Eq. (9) with varying  $M$  across rows. Starting from the top row, the observer must correctly identify at least half of the optotypes on a row before proceeding to the next row. If they can't identify the optotype on the top row, the VA score was 1.1  $\log\text{MAR}$ ; otherwise, the VA score is equal to the size of the optotypes in the last row with at least 50% correct identification. The VA score could therefore take 12 potential values. For each simulated observer  $x_i$ , we repeated the test 1000 times to obtain the distribution of test scores  $P(y_j|x_i)$ , where  $j = 1, \dots, 12$ .

#### ETDRS chart

The ETDRS chart (Fig. 3b) has 14 five-optotype rows, with optotype size descending from 1.0 to  $-0.3 \log\text{MAR}$ . Each simulated observer was tested with the standard procedure<sup>26,50</sup>. The probability of correctly identifying  $m$  optotypes in a row is determined by Eq. (9) with  $M = 5$ . Four different termination rules were simulated. Starting from the top row, the test could stop after the observer makes three, four, or five mistakes in identifying the optotypes in a row or continue until the observer is tested with the entire chart. The VA score is computed as  $1.1 - 0.02n$ , where  $n$  is the number of correctly identified optotypes, with 71 potential values. For each simulated observer  $x_i$ , we repeated the test 1000 times to obtain the distribution of test scores  $P(y_j|x_i)$ , where  $j = 1, \dots, 71$ .

#### qVA test

The qVA (Fig. 3c) is a Bayesian active learning visual acuity test<sup>27</sup>. Its stimulus space consists of optotypes of 91 linearly spaced sizes from  $-0.5$  to  $1.3 \log\text{MAR}$ , with a  $0.02 \log\text{MAR}$  resolution. Starting with a weak prior distribution of VA threshold and VA range in a two-dimensional space that has 700 linearly spaced VA thresholds (between  $-0.5$  and  $1.3 \log\text{MAR}$ ) and 699 log-linearly spaced VA ranges (between  $0.1$  and  $1.5 \log\text{MAR}$ ), it uses an active learning procedure to test the observer with the optimal stimulus in each trial and generates the posterior distribution of VA threshold and range. Each simulated observer was tested with 5, 15, or 30 rows (corresponding to 15, 45, or 90 optotypes). The probability of correctly identifying  $m$  optotypes in a row is determined by Eq. (9) with  $M = 3$ . We computed the mean VA threshold and range from their posterior distributions in each test and quantized them into 86 and 56 discrete scores with a  $0.02 \log\text{MAR}$  resolution, with a total of 4816 potential combinations. For each simulated observer  $x_i$ , we repeated the test 1000 times to obtain the two-dimensional distribution of test scores  $P(y_j|x_i)$ , with  $j = 1, \dots, 4816$ . We also computed the distribution of VA threshold  $P(y_{\text{Threshold},j}|x_i)$  by marginalizing  $P(y_j|x_i)$ :

$$P(y_{\text{Threshold},j}|x_i) = \sum_{\text{range}} P(y_j|x_i), \quad (10)$$

where  $j = 1, \dots, 86$ .

#### Information gain

We first computed  $P(y_j)$  from  $P(y_j|x_i)$  for each test:

$$P(y_j) = \frac{1}{I} \sum_{i=1}^I P(y_j|x_i), \quad (11)$$

where  $I = 1386$  in both simulations. We then computed the total entropy of each test of a population:

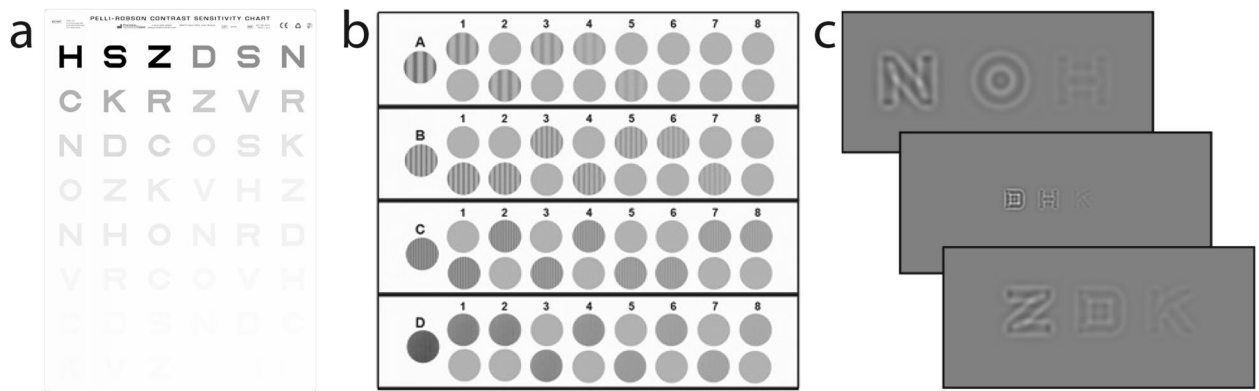
$$H(Y) = - \sum_{j=1}^J P(y_j) \log_2(P(y_j)), \quad (12)$$

where  $J = 12, 71, 86, 4816$  for the Snellen, ETDRS, VA threshold from qVA, and VA threshold and VA range from qVA. The expected residual entropy was computed as:

$$H(Y|X) = - \frac{1}{I} \sum_{i=1}^I \sum_{j=1}^J P(y_j|x_i) \log_2(P(y_j|x_i)), \quad (13)$$

and, finally expected information gain can be obtained:

$$IG(Y|X) = H(Y) - H(Y|X). \quad (14)$$



**Figure 4.** (a) A Pelli–Robson chart (Image courtesy of Precision Vision, Inc.). (b) A CSV-1000 chart (Reproduced with permission from GUARDiON Health Sciences, Inc.). (c) A subset of potential stimuli in qCSF.

## Contrast sensitivity tests

### Simulated observers

We conducted two simulations using the Pelli–Robson chart, CSV-1000 chart, and qCSF test (Fig. 4). In Simulation 1, we simulated 1911 observers from a uniform distribution of peak gain ( $\theta_{PG}^{CSF}$ ), peak spatial frequency ( $\theta_{PF}^{CSF}$ ), and band width ( $\theta_{BH}^{CSF}$ ), with  $\log_{10}(\theta_{PG}^{CSF}) \in [0.3, 2.3]$  sampled every 0.10  $\log_{10}$  units,  $\log_{10}(\theta_{PF}^{CSF}) \in [-0.3, 0.9]$  sampled every 0.10  $\log_{10}$  units, and  $\log_{10}(\theta_{BH}^{CSF}) \in [0.3, 0.6]$  sampled every 0.05  $\log_{10}$  units. In Simulation 2, we simulated 1911 observers from the population distribution of CSF parameters derived from two existing qCSF datasets, one consisted of 112 eyes tested in three luminance conditions<sup>38</sup> and the other of 14 eyes tested with Bangerter foils<sup>28</sup>, using a hierarchical Bayesian model<sup>49,51</sup>. The original experiments were conducted at the Ohio State University. Written consent was obtained from all the participants before the experiment. The study protocol was approved by the institutional review board of human subject research of The Ohio State University and adhered to the tenets of the Declaration of Helsinki.

### Letter and grating contrast sensitivity functions

The letter CSF, which specifies contrast sensitivity  $S_{letter}(f)$  for filtered letters of different sizes at center spatial frequency  $f$ , can be described with a log parabola function with three parameters  $\theta^{CSF} = (\theta_{PG}^{CSF}, \theta_{PF}^{CSF}, \theta_{BH}^{CSF})$ <sup>20,52,53</sup>:

$$\log_{10}(S_{letter}(f|\theta^{CSF})) = \log_{10}(\theta_{PG}^{CSF}) - \frac{4}{\log_{10}(2)} \left( \frac{\log_{10}(f) - \log_{10}(\theta_{PF}^{CSF})}{\theta_{BH}^{CSF}} \right)^2, \quad (15)$$

where  $\theta_{PG}^{CSF}$  is the peak gain,  $\theta_{PF}^{CSF}$  is the peak spatial frequency (cycles/degree), and  $\theta_{BH}^{CSF}$  is the bandwidth (octaves) at half-height. For an observer with peak gain  $\theta_{PG}^{CSF}$ , peak spatial frequency  $\theta_{PF}^{CSF}$ , and bandwidth  $\theta_{BH}^{CSF}$ , the probability of correct identification of a bandpass-filtered optotype with contrast  $c$  and center spatial frequency  $f$  is described with a Weibull psychometric function<sup>34</sup>:

$$p(\text{correct}|\theta^{CSF}, f, c) = g + \left(1 - g - \frac{\lambda}{2}\right) [1 - \exp[-(c \times S_{letter}(f|\theta^{CSF}))^b]], \quad (16)$$

where  $g$  is the guessing rate,  $\lambda = 0.04$  is the lapse rate, and  $b$  determines the steepness of the psychometric function. Because they both use a 10-alternative forced identification task,  $g = 0.1$  and  $b = 4.05$  for the Pelli–Robson chart and qCSF test.

The grating CSF, which specifies contrast sensitivity  $S_{grating}(f)$  for gratings of the same size at spatial frequency  $f$ , needs to be corrected for the increased number of cycles with increasing spatial frequency<sup>54</sup>. For the grating stimuli used in the CSV-1000:

$$\log_{10}(S_{grating}(f|\theta^{CSF})) = \log_{10}(\theta_{PG}^{CSF}) - \frac{4}{\log_{10}(2)} \left( \frac{\log_{10}(f) - \log_{10}(\theta_{PF}^{CSF})}{\theta_{BH}^{CSF}} \right)^2 + \log_{10}(0.539f). \quad (17)$$

For the yes/no task in the first column of the CSV-1000 test, we used a high-threshold model. That is, the simulated observer says yes if the stimulus contrast > threshold ( $= 1/S_{grating}(f|\theta^{CSF})$ ). The simulated observer says no otherwise. For the two-alternative forced choice task in CSV-1000, we replace  $S_{letter}(f|\theta^{CSF})$  with  $S_{grating}(f|\theta^{CSF})$  and set  $g = 0.5$ ,  $b = 3.06$  in Eq. (16) to compute the probability of making a correct response.

### Pelli–Robson chart

The Pelli–Robson chart (Fig. 4a) consists of 16 optotype triplets of the same size and log-linearly spaced contrast between 0.56 and 100%<sup>33</sup>. At a viewing distance of 3 m, the center frequency of the optotypes is 3 c/d. The

probability of correctly identifying each optotype in the chart is determined by Eq. (16). Starting from the top row, the test proceeds to the next triplets only if the observer correctly identifies at least two of the three optotypes in the current triplet. The CS of the observer is determined by the lowest contrast  $c_{lowest}$  at which they correctly identify at least two of the three letters in the triplet:  $S_{letter}(3c/d) = -\log_{10}(c_{lowest})$ , with 17 potential contrast sensitivity scores. For each simulated observer  $x_i$ , we repeated the test 1000 times to obtain the distribution of test scores  $P(y_j|x_i)$ , where  $j = 1, \dots, 17$ .

#### CSV-1000 chart

The CSV-1000 chart (Fig. 4b) consists of CS tests at four spatial frequencies: 3, 6, 12, and 18 cycles/degree. Each test consists of 17 stimuli arranged in nine columns, with a single high-contrast vertical sinewave grating in the first column, and two test patches in the remaining eight columns, of which only one contains a vertical sinewave grating. The gratings are arranged with decreasing contrast from left to right, with contrast from  $-0.70$  to  $-2.08$ ,  $-0.91$  to  $-2.29$ ,  $-0.61$  to  $-1.99$ , and  $-0.17$  to  $-1.55$  log10 units in the four rows. Going through all four rows starting from the top, the observer is first required to perform a yes/no task on the first column in each row. If the observer can't see the stimulus in the first column, the test stops for that row and the observer's contrast sensitivity is:

$$S(f) = \begin{cases} -\log_{10}(c_{firstcolumn}(f)) - 0.3, & f = 3, 6, 12 \\ 0.01, & f = 18 \end{cases} \quad (18)$$

If the observer can see the stimulus in the first column, they proceed to identify the location of the patch that contained the grating in each column with a three-alternative forced choice response: top, bottom, or blank. We treat blank as an incorrect response. The lowest contrast at which the observer correctly identifies the location of the grating is used to determine CS in the test:  $S(f) = -\log_{10}(c_{lowest}(f))$ . The result is a four-dimensional CS score sampled at four spatial frequencies. Because there are 10 potential CS scores in each spatial frequency, there are therefore a total of  $10^4$  potential CS functions. For each simulated observer  $x_i$ , we repeated the test 1000 times to obtain the distribution of test scores  $P(y_j|x_i)$ , where  $j = 1, \dots, 10^4$ .

#### qCSF test

The qCSF (Fig. 4c) is a Bayesian active learning contrast sensitivity test<sup>20,34</sup>. Its stimulus space consists of 128 log-linearly spaced contrasts (from 0.002 to 1.0) and 19 log-linearly spaced spatial frequencies (from 1.19 to 30.95 c/d). Although a four-parameter truncated log parabola has been used in the qCSF test<sup>20</sup>, we removed the truncation parameter in the simulations because we didn't score the simulated observers in very low spatial frequencies. Starting with a weak prior distribution of peak gain, peak frequency and bandwidth in a three-dimensional space that has 60 log-linearly spaced peak gains (from 1.05 to 1050), 40 log-linearly spaced peak frequencies (from 0.1 to 20 c/d), and 27 log-linearly spaced bandwidth (from 1 to 9 octaves), it uses an active learning procedure to test the observer with the optimal stimuli in each trial and computes the posterior distribution of the three CSF parameters in the afore-mentioned three-dimensional space. In each trial, three filtered optotypes with the same center spatial frequency but four, two, and one times the optimal contrast (capped at 0.9) are presented. The observer could be tested with 15, 25 and 50 trials. For each simulated observer  $x_i$ , we repeated the test 1000 times to obtain distributions of the unidimensional AULCSF  $P(y_{AULCSF,j}|x_i)$  and the six-dimensional CSF score  $P(y_{CSF,j}|x_i)$  at six spatial frequencies (1, 1.5, 3, 6, 12 and 18 c/d). Sampling the scores at 0.05 log10 resolution,  $j = 1, \dots, 57$  for  $y_{AULCSF,j}$ , and  $j = 1, \dots, 253,492$  for  $y_{CSF,j}$ .

#### Information gain

Equations (11)–(14) were used to compute  $P(y_j)$ ,  $H(Y)$ ,  $H(Y|X)$ , and  $IG(Y|X)$ , with  $I = 1911$  for the two simulations, and  $J = 17, 10,000, 57, 253,492$  for the Pelli–Robson, CSV-1000, AULCSF from qCSE, and CSF from qCSE.

## Results

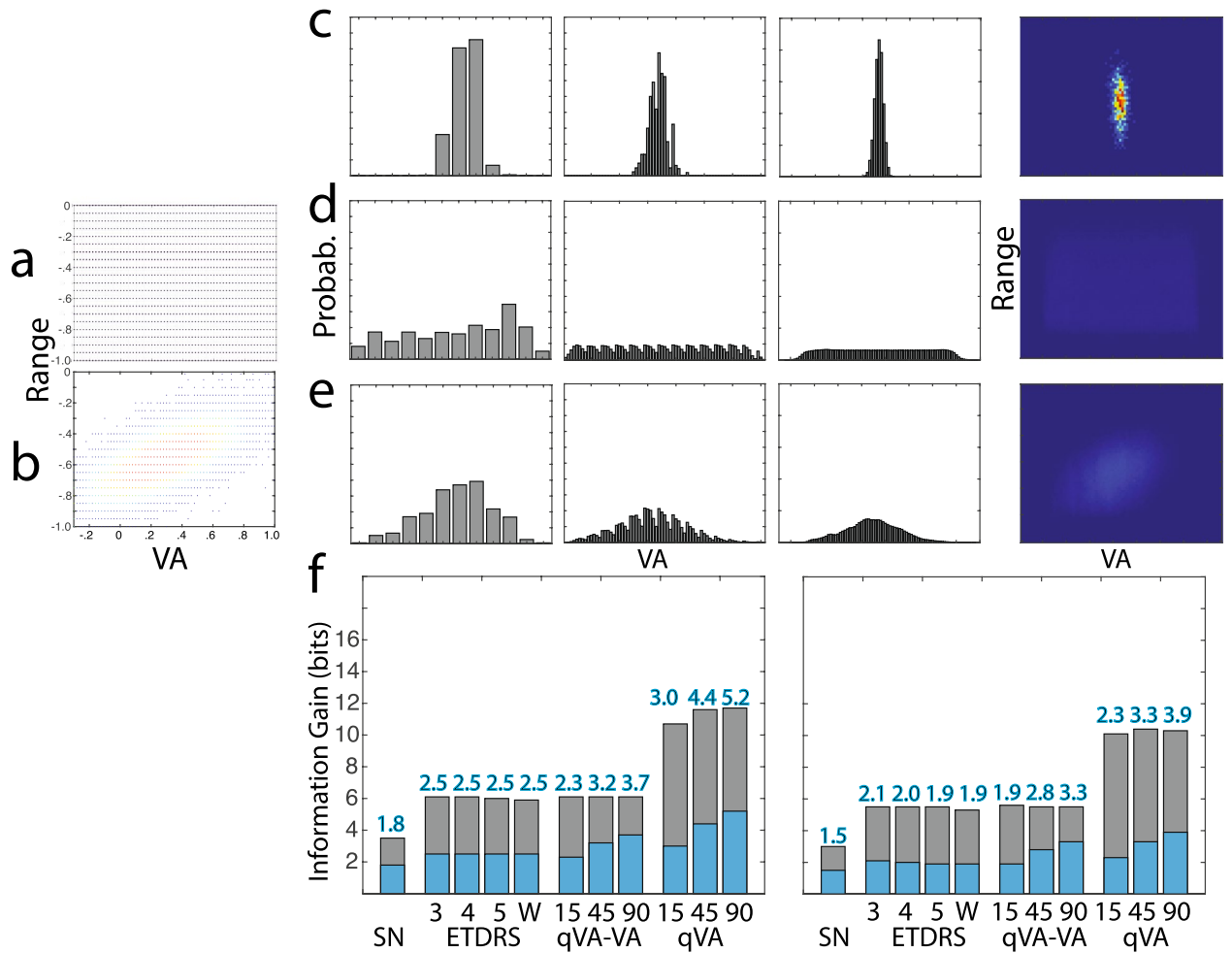
### Visual acuity tests

The VA threshold and VA range distributions of the observers in the two simulations are shown in Fig. 5a,b. Figure 5c shows distributions of the test scores  $P(y_j|x_i)$  of one representative simulated observer  $x_i$  in the Snellen, ETDRS (3-mistake rule), and qVA test, with results from the qVA test scored as VA-alone, and as both VA threshold and VA range. Figure 5d,e show the distributions of the test scores  $P(y_j)$  of the populations in Simulations 1 and 2, respectively. Because a uniform  $X$  distribution is used in Simulation 1, the corresponding  $P(y_j)$ 's from the tests are nearly uniform. On the other hand,  $P(y_j)$ 's in Simulation 2 are more concentrated because the population is more concentrated.

The total entropy  $H(Y)$ , the expected residual entropy  $H(Y|X)$ , the expected information gain  $IG(Y|X)$ , the expected number of classes  $N(Y|X)$ , the average number of optotypes tested, the expected information gain per optotype tested, and the ratio of expected information gain versus the  $\log_2$  of the number of optotypes tested from the three tests in the two simulations are listed in Table 1. The expected information gain  $IG(Y|X)$  is also plotted in Fig. 5f. As expected,  $H(Y)$  in Simulation 2 is less than the corresponding  $H(Y)$  in Simulation 1 for all the tests because of the more concentrated  $P(y_j)$  resulted from a narrower range of simulated observers. As a result,  $IG(Y|X)$  and  $N(Y|X)$  in Simulation 2 are less than those in Simulation 1.

To check the validity of the simulations, we also estimated expected information  $IG(Y|X)$  of the Snellen and ETDRS from their reported test–retest variabilities ( $TRV = 1.96\sigma$ ), with the assumption that the outcome distributions are normal and have the same TRV for observers with different acuities. For both tests, the outcome scores cover  $-0.3$  to  $1.0$  logMAR, expected information gain can be computed with  $IG(Y|X) = \log_2\left(\frac{L}{4.13\sigma}\right)$ , where





**Figure 5.** Distributions and information gain in visual acuity tests. (a,b) Distributions of VA threshold and VA range of the simulated observers in Simulations 1 and 2. (c) From left to right: Distributions of the test scores  $P(y_j|x_i)$  of one representative simulated observer  $x_i$  in the Snellen, ETDRS (3-mistake rule), and qVA test (45 optotypes), with results from the qVA test scored as VA-alone, and as both VA threshold and VA range. (d,e) Distributions of the test scores  $P(y_j)$  of the populations in Simulations 1 and 2, respectively. (f) Expected information gain  $IG(Y|X)$  of the various tests in Simulation 1 (left) and Simulation 2 (right). For ETDRS, results from the 3-, 4-, 5-mistake and whole chart rules are shown. For qVA, results from testing with 15, 45 and 90 optotypes are shown. The blue bars represent  $IG(Y|X)$ ; the gray bars represent residual entropy  $H(Y|X)$ .

$L = 1.3 \log\text{MAR}$ . For the Snellen chart, the typical TRV is  $0.23 \log\text{MAR}^{35}$ , with  $\sigma = \frac{\text{TRV}}{1.96} = 0.117 \log\text{MAR}$ , and the expected information gain is 1.4 bits. For the ETDRS chart, the typical TRV is  $0.11 \log\text{MAR}^{35}$ , with  $\sigma = \frac{\text{TRV}}{1.96} = 0.056 \log\text{MAR}$ , and the expected information gain is 2.5 bits. These estimated values are largely consistent with our results in Simulation 1.

For the qVA test, we also computed expected information gain from an existing dataset with 14 eyes tested in four Bangerter foil conditions<sup>28</sup>. In this dataset, VA threshold is between  $-0.15$  and  $0.68 \log\text{MAR}$ , and VA range is between  $0.12$  and  $0.63 \log\text{MAR}$ . We computed the total entropy, residual entropy and information gain based on the posterior distributions from single tests rather than distributions of test scores from repeated tests. In the qVA test, the posterior distributions from single tests are broader than those derived from repeated tests until about 45 optotypes are tested and converge to those from repeated tests afterwards<sup>28</sup>. The expected information gain based on VA threshold alone is 1.6, 2.6 and 2.9 bits after qVA test with 15, 45 and 90 optotypes, respectively, and the expected information gain based on VA threshold and range is 2.1, 3.2, and 3.6 bits after qVA test with 15, 45 and 90 optotypes, respectively. These results are largely consistent with those from Simulation 2.

In both simulations, ETDRS generated more expected information gain than Snellen. Scored with VA threshold alone or with both VA threshold and VA range, qVA with 15 rows (or 45 optotypes) generated more expected information gain than ETDRS. In terms of expected information gain per optotype tested, the qVA test with 45 optotypes scored with VA alone and with both VA and VA range was more efficient than the Snellen chart, which was in turn more efficient than the ETDRS chart. The different efficiencies were caused by different test designs and the distributions of acuity behavior used in the study. Interestingly, the ratio of expected information gain and the  $\log_2$  of the number of optotypes tested in qVA was essentially constant across test lengths and was

		Snellen	ETDRS				Whole chart	qVA					
			# Mistakes			15		VA alone			VA + Range		
			3	4	5			45	90	15	45	90	
Simulation 1	H(Y) (bits)	3.5	6.1	6.1	6.0	5.9	6.1	6.1	6.1	10.7	11.6	11.7	
	H(Y X) (bits)	1.6	3.5	3.5	3.6	3.4	3.8	2.9	2.4	7.8	7.2	6.5	
	IG (bits)	1.8	2.5	2.5	2.5	2.5	2.3	3.2	3.7	3.0	4.4	5.2	
	N(Y X)	3.6	5.7	5.7	5.6	5.6	5.0	9.1	12.7	7.9	20.5	37.2	
	# Optotypes	24.7	41.4	45.5	51.8	70	15	45	90	15	45	90	
	IG/Optotype(bits)	0.07	0.06	0.06	0.05	0.04	0.16	0.07	0.04	0.20	0.10	0.06	
	IG/log2(# Optotype)(bits)	0.40	0.47	0.46	0.44	0.40	0.60	0.58	0.56	0.76	0.79	0.80	
Simulation 2	H(Y) (bits)	3.0	5.5	5.5	5.5	5.3	5.6	5.5	5.5	10.1	10.4	10.3	
	H(Y X) (bits)	1.6	3.5	3.5	3.6	3.4	3.7	2.8	2.3	7.7	7.2	6.5	
	IG (bits)	1.5	2.1	2.0	1.9	1.9	1.9	2.8	3.3	2.3	3.3	3.9	
	N(Y X)	2.8	4.2	4.0	3.8	3.8	3.8	6.8	9.5	5.1	9.6	14.6	
	# Optotypes	23.8	44.7	48.4	55.1	70	15	45	90	15	45	90	
	IG/Optotype(bits)	0.06	0.05	0.04	0.03	0.03	0.13	0.06	0.04	0.16	0.07	0.04	
	IG/log2(# Optotype)(bits)	0.32	0.38	0.36	0.33	0.32	0.49	0.50	0.50	0.60	0.59	0.60	

**Table 1.** Entropy, information gain and number of classes from visual acuity tests.

greater than the corresponding scores in ETDRS and Snellen. The results suggest that the expected information gain increased linearly with the  $\log_2$  of the number of optotypes in the qVA test.

### Contrast sensitivity tests

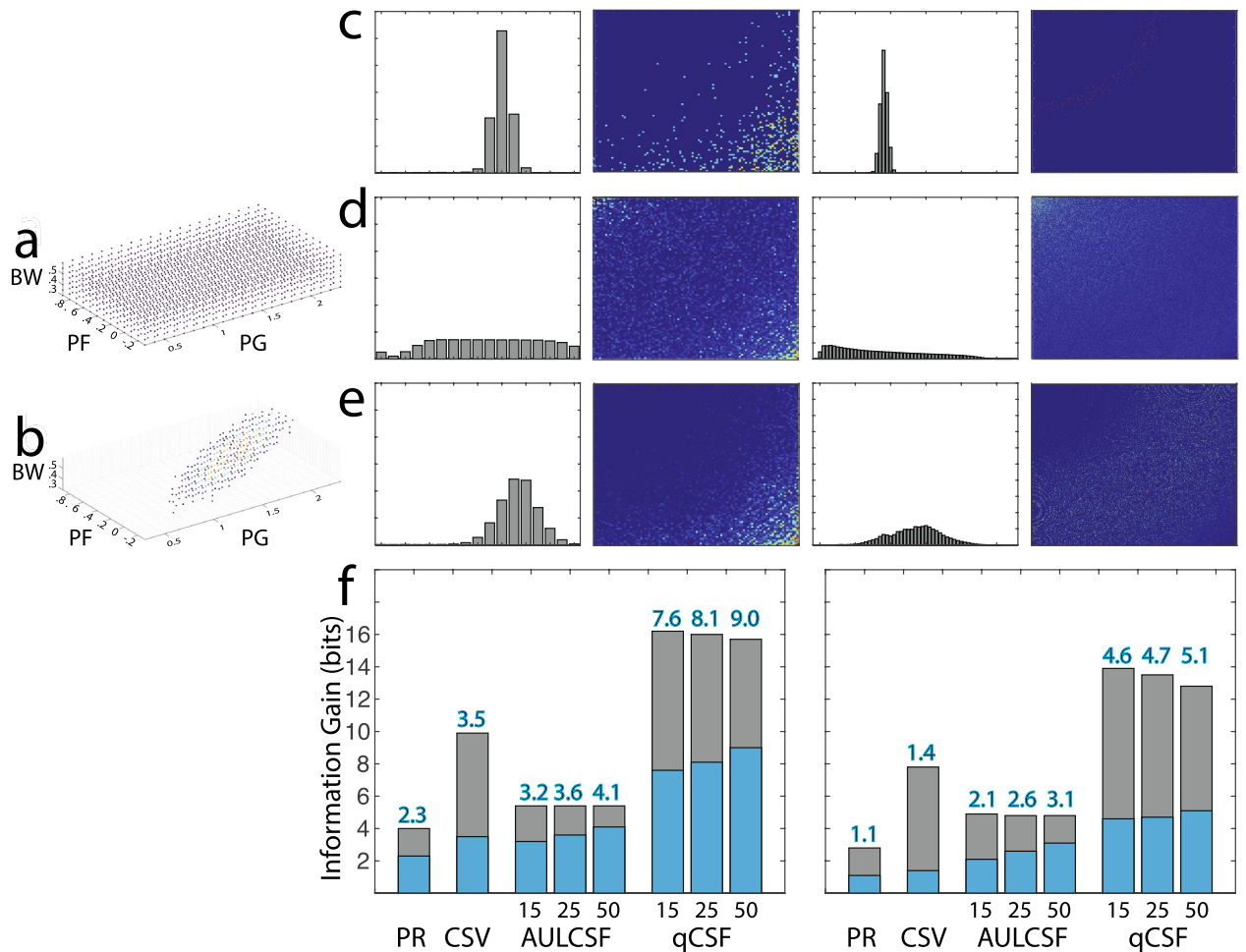
The peak gain, peak spatial frequency, and bandwidth distributions of the observers in the two simulations are shown in Fig. 6a,b. Figure 6c shows distributions of the test scores  $P(y_j|x_i)$  of one representative simulated observer  $x_i$  in the Pelli–Robson, CSV-1000, and qCSF test, with results from the qCSF test scored as AULCSF, and CSF at six spatial frequencies. Figure 6d,e show the distributions of the test scores  $P(y_j)$  of the populations  $X$  in Simulations 1 and 2, respectively. Because a uniform  $X$  distribution is used in Simulation 1, the corresponding  $P(y_j)$ s from the tests are more uniform. On the other hand,  $P(y_j)$ s in Simulation 2 are more concentrated because the observers are sampled in a narrower range.

The total entropy  $H(Y)$ , the expected residual entropy  $H(Y|X)$ , the expected information gain  $IG(Y|X)$ , the expected number of classes  $N(Y|X)$ , the average number of optotypes tested, the expected information gain per optotype tested, and the ratio of expected information gain versus the  $\log_2$  of the number of optotypes tested from the three tests in the two simulations are listed in Table 2. The expected information gain  $IG(Y|X)$  is also plotted in Fig. 6f. As expected,  $H(Y)$  in Simulation 2 is less than the corresponding  $H(Y)$  in Simulation 1 in all the tests because of the more concentrated  $P(y_j)$  resulted from a narrower range of simulated observers. As a result,  $IG(Y|X)$  and  $N(Y|X)$  in Simulation 2 are less than those in Simulation 1.

To check the validity of the simulations, we also estimated information  $IG(Y|X)$  of the Pelli–Robson test from its reported test–retest variabilities ( $TRV = 1.96\sigma$ ), with the assumption that the outcome distributions are normal and have the same TRV for all the observers. For the test, the outcome scores cover 0 to  $2.25 \log_{10}$  contrast sensitivity, with a typical TRV between 0.15 and  $0.20 \log_{10}$  in the normal population<sup>55</sup>. Therefore,  $L = 2.25 \log_{10} CS$ ,  $\sigma = \frac{TRV}{1.96}$  is between 0.077 and  $0.10 \log_{10} CS$ , and the expected information gain is between 2.4 and 2.8 bits. These estimated values are largely consistent with the results in Simulation 1.

For the qCSF test, we also computed information gain from two existing datasets, one with 112 subjects tested binocularly in three luminance conditions<sup>38</sup> and the other with 14 eyes tested monocularly in four Bangerter foil conditions<sup>28</sup>. In this dataset, peak gain is between 0.92 and  $2.27 \log_{10} CS$ , peak spatial frequency is between 0.21 and  $3.9 c/d$ , and the bandwidth is between 1.8 and 5.7 octaves. We computed the total entropy, expected residual entropy, and expected information gain based on the posterior distributions from single tests rather than distributions of test scores from repeated tests. In the qCSF test, the posterior distributions from single tests are broader than those derived from repeated tests until about 25 trials are tested and converge to those from repeated tests afterwards<sup>34</sup>. The expected information gain based on AULCSF is 2.0, 2.4, and 2.9 bits after qCSF test with 15, 25 and 50 trials, respectively, and the expected information gain based on CSF at six spatial frequencies is 4.1, 4.6, and 5.3 bits after qCSF test with 15, 25 and 50 trials, respectively. Again, the results from the dataset are largely consistent with those from Simulations 2.

In both simulations, CSV-1000 generated more expected information gain than the Pelli–Robson test. Scored with AULCSF or with CSF at six spatial frequencies, qCSF with 25 trials generated more expected information gain than CSV-1000. In terms of expected information gain per optotype tested, the CSV-1000 test was the most efficient in Simulation 1, and qCSF with 15 rows of trials was the most efficient in Simulation 2. The variations in efficiency were caused by different test designs and the distributions of the populations used in the study. Interestingly, the ratio of expected information gain versus the  $\log_2$  of the number of optotypes tested in qCSF was essentially constant across test lengths and was greater than the corresponding scores in Pelli–Robson in



**Figure 6.** Distributions and information gain in contrast sensitivity tests. **(a,b)** Distributions of peak gain, peak spatial frequency, and bandwidth of the observers in Simulations 1 and 2. **(c)** From left to right: Distributions of the test scores  $P(y_j|x_i)$  of one representative simulated observer  $x_i$  in the Pelli–Robson, CSV-1000, and qCSF test (25 trials), with results from the qCSF test scored as AULCSF, and CSF at six spatial frequencies. **(d,e)** Distributions of the test scores  $P(y_j)$  of the populations in Simulations 1 and 2, respectively. **(f)** Expected information gain  $IG(Y|X)$  of the various tests in Simulation 1 (left) and Simulation 2 (right). For qCSF, results from testing with 15, 25, and 50 trials are shown. The blue bars represent  $IG(Y|X)$ ; the gray bars represent residual entropy  $H(Y|X)$ .

Simulation 1 and both Pelli–Robson and CSV-1000 in Simulation 2. The results suggest that the expected information gain increased linearly with the  $\log_2$  of the number of optotypes in the qCSF test.

## Discussion

In this study, we introduced a concept from information theory, called expected information gain (or mutual information), to quantify the amount of new knowledge that can be obtained from measurement on a population. This concept allows us to compare measurements with different dimensionalities and assess the potential advantages of new measurements that generate distinct or higher dimensional data compared to the current gold standard. Importantly, it focuses on the actual knowledge gained through the measurements, surpassing the mere quantification of measurement uncertainties.

We demonstrated two equivalent approaches for computing expected information gain:

- (1) *Reduction of uncertainty* This approach gauges the reduction in uncertainty regarding the “truth” (the property to be measured) after the measurement.
- (2) *Difference in uncertainty* The second approach calculates the difference between the uncertainty associated with all possible measurement outcomes in a population and the expected residual uncertainty after the measurement.

In both approaches, the key idea is that expected information gain quantifies the increase in knowledge achieved through measurement. This knowledge gain is greater when there is more uncertainty initially and/or

		Pelli–Robson	CSV-1000	qCSF					
				AULCSF			CSF		
				15	25	50	15	25	50
Simulation 1	H(Y) (bits)	4.0	9.9	5.4	5.4	5.4	16.2	16.0	15.7
	H(Y X) (bits)	1.7	6.4	2.2	1.8	1.4	8.5	7.8	6.7
	IG (bits)	2.3	3.5	3.2	3.6	4.1	7.6	8.1	9.0
	Categories	5.0	11.1	9.1	12.1	16.7	196.1	281.2	498.0
	# Optotypes	27.9	11.0	45	75	150	45	75	150
	IG/Optotype(bits)	0.08	0.32	0.07	0.05	0.03	0.17	0.11	0.06
	IG/log2(# Optotype)(bits)	0.48	1.01	0.58	0.58	0.56	1.39	1.31	1.24
Simulation 2	H(Y) (bits)	2.8	7.8	4.9	4.8	4.8	13.9	13.5	12.8
	H(Y X) (bits)	1.7	6.5	2.7	2.3	1.7	9.4	8.8	7.8
	IG (bits)	1.1	1.4	2.1	2.6	3.1	4.6	4.7	5.1
	Categories	2.1	2.6	4.3	5.9	8.6	23.7	26.0	34.1
	# Optotypes	36.8	15.2	45	75	150	45	75	150
	IG/Optotype(bits)	0.03	0.09	0.05	0.03	0.02	0.10	0.06	0.03
	IG/log2(# Optotype)(bits)	0.21	0.35	0.38	0.41	0.43	0.83	0.75	0.70

**Table 2.** Entropy, information gain and number of categories from contrast sensitivity tests.

when the measurement effectively reduces this uncertainty. These approaches provide a rigorous and quantitative way to assess the value of measurement in various fields of science and beyond.

As a practical application of this concept, we computed and compared expected information gain from various VA and CS tests. Our findings revealed some intriguing insights:

- (1) *VA tests* The ETDRS chart, the current gold standard, generated less expected information gain than the qVA test with 15 rows, indicating the potential for improved VA assessment with qVA.
- (2) *CS tests* The CSV-1000 test outperformed the Pelli–Robson test, but the qCSF test with 25 trials demonstrated the highest expected information gain, showcasing its advantages in assessing CS.

It's crucial to consider the target population when comparing expected information gain, as it can vary for different populations. Our simulations demonstrated this dependence, indicating that different target populations might require different measurements, and different measurements may have advantages or disadvantages in different populations. Therefore, when comparing tests, it is crucial to consider the target population alongside the measurement instruments. In this study, we used the same target populations to compare different VA and CS tests, and consistent rank orders of the tests were obtained across the two populations.

Expected information gain isn't limited to vision tests; it can be applied across various domains. For example, it can help compare newly developed optical coherence tomography (OCT) tests with existing ones, assess different vision tests' abilities to classify patients, or evaluate data analytics techniques based on their capacity to reduce uncertainty<sup>49,51</sup>.

Expected information gain can serve as a valuable tool for objectively evaluating the new knowledge acquired through measurements. It may facilitate better comparisons and informed decision-making in diverse fields of research and data analysis.

### Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

### Appendix 1: Proof of equivalence

It's well-known that information gain is symmetric, that is,  $IG(Y|X) = IG(X|Y)$ . We can prove this by substituting  $P(x|y)$  with its Bayesian computation in Eq. (2):

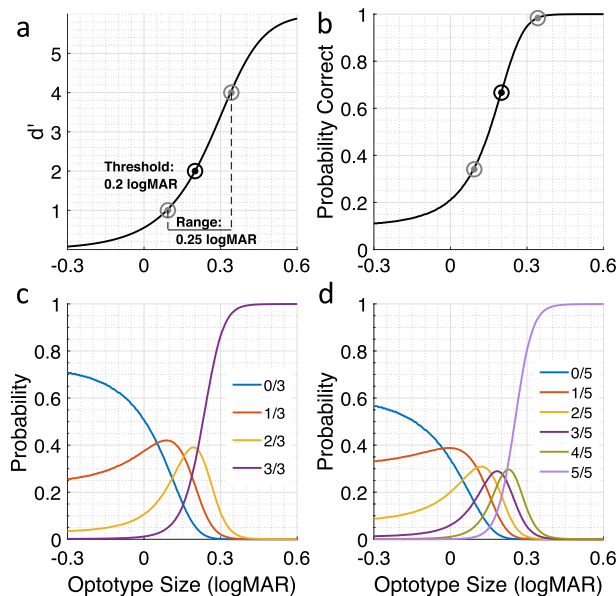
$$\begin{aligned}
 H(X|Y) &= - \int_Y \left( \int_X (P(x|y) \log_2(P(x|y))) dx \right) P(y) dy \\
 &= - \int_Y \left( \int_X \frac{P(y|x)P(x)}{P(y)} \log_2 \left( \frac{P(y|x)P(x)}{P(y)} \right) dx \right) P(y) dy \\
 &= \int_Y \left( \int_X P(y|x)P(x) \log_2(P(y)) dx \right) dy - \int_Y \left( \int_X P(y|x)P(x) \log_2(P(y|x)P(x)) dx \right) dy \\
 &= \int_Y P(y) \log_2(P(y)) dy - \int_X \left( \int_Y P(y|x) \log_2(P(y|x)) dy \right) P(x) dx - \int_X P(x) \log_2(P(x)) dx.
 \end{aligned}
 \tag{A1}$$

Therefore,

$$IG(X|Y) = H(X) - H(X|Y) = - \int_Y P(y) \log_2(P(y)) dy + \int_X \left( \int_Y P(y|x) \log_2(P(y|x)) dy \right) P(x) dx = IG(Y|X).
 \tag{A2}$$

### Appendix 2: Multi-optotype acuity behavioral psychometric functions

From the  $\log(d')$  acuity behavioral psychometric function for a single-optotype in Eq. (8) (Fig. 7a), we can derive the percent correct psychometric function for a single optotype for the N-alternative forced optotype identification (N-AFC) task based on signal detection theory (Fig. 7b), with N = 10 in the Snellen, ETDRS and qVA tests<sup>21</sup>. We then take into account of the chart design (i.e., optotype sampling with or without replacement) to derive the probability of correctly identifying different numbers optotypes in each row of a test. The multi-optotype psychometric functions are shown for the qVA (Fig. 7c) and ETDRS (Fig. 7d) tests. Specially, the three optotypes in each test row are sampled from 10 Sloan letters with replacement in the qVA test, while the five optotypes in each test row are sampled from 10 Sloan letters with replacement in the ETDRS test.



**Figure 7.** Multi-optotype acuity behavioral psychometric functions. **(a)** The  $d'$  psychometric function for a single optotype for an observer with VA threshold = 0.2 logMAR and VA range = 0.25 logMAR. **(b)** The percent correct psychometric function for single-optotype in a 10-AFC task for the observer in **(a)**. **(c)** The multi-optotype psychometric functions of the observer in the qVA test, i.e., the probabilities of correctly identifying 0, 1, 2, or 3 of the three optotypes in a test row as functions of optotype size of the row. **(d)** The multi-optotype psychometric functions of the observer in the ETDRS test, the probabilities of correctly identifying 0, 1, 2, 3, 4 or 5 of the five optotypes presented in a test row as functions of optotype size of the row.

Received: 6 June 2023; Accepted: 29 September 2023

Published online: 05 October 2023

## References

1. Newton, I. *Philosophiæ Naturalis Principia Mathematica* Vol. 1 (G. Brookman, 1833).
2. Ellis, B. *Basic Concepts of Measurement* (1968).
3. Gauch, H. G. & Gauch, H. G. *Scientific Method in Practice* (Cambridge University Press, 2003).
4. Gower, B. *Scientific Method: A Historical and Philosophical Introduction* (Routledge, 2012).
5. Penrose, R. *The Road to Reality: A Complete Guide to the Laws of the Universe* (Random House, 2005).
6. Houle, D., Pélabon, C., Wagner, G. P. & Hansen, T. F. Measurement and meaning in biology. *Q. Rev. Biol.* **86**(1), 3–34 (2011).
7. Krantz, D., Luce, D., Suppes, P. & Tversky, A. *Foundations of Measurement, Vol. I: Additive and Polynomial Representations* (1971).
8. Liptak, B. G. *Instrument Engineers' Handbook, Volume One: Process Measurement and Analysis* (CRC Press, 2003).
9. De Vet, H. C., Terwee, C. B., Mokkink, L. B. & Knol, D. L. *Measurement in Medicine: A Practical Guide* (Cambridge University Press, 2011).
10. Standardization IOF. *Accuracy (Trueness and Precision) of Measurement Methods and Results—Part 1: General Principles and Definitions* (1994).
11. International Organization for Standardization. *Precision of Test Methods—Determination of Repeatability and Reproducibility for a Standard Test* (International Organization for Standardization, 1986).
12. Bland, J. M. & Altman, D. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* **327**(8476), 307–310 (1986).
13. Nyquist, H. Certain topics in telegraph transmission theory. *Trans. Am. Inst. Electr. Eng.* **47**(2), 617–644 (1928).
14. Hartley, R. V. Transmission of information 1. *Bell Syst. Tech. J.* **7**(3), 535–563 (1928).
15. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**(3), 379–423 (1948).
16. Douglas, H. *How to Measure Anything: Finding the Value of Intangibles in Business* (Wiley, 2007).
17. Cover, T. M. & Thomas, J. A. *Information Theory and the Stock Market. Elements of Information Theory* 543–556 (Wiley, 1991).
18. Kullback, S. & Leibler, R. A. On information and sufficiency. *Ann. Math. Stat.* **22**(1), 79–86 (1951).
19. Larose, D. T. & Larose, C. D. *Discovering Knowledge in Data: An Introduction to Data Mining* Vol. 4 (Wiley, 2014).
20. Lesmes, L. A., Lu, Z.-L., Baek, J. & Albright, T. D. Bayesian adaptive estimation of the contrast sensitivity function: The quick CSF method. *J. Vis.* **10**(3), 17 (2010).
21. Lu, Z.-L. & Doshier, B. *Visual Psychophysics: From Laboratory to Theory* (MIT Press, 2013).
22. Kim, W. *et al.* A hierarchical adaptive approach to optimal experimental design. *Neural Comput.* **26**(11), 2465–2492 (2014).
23. Kniestedt, C. & Stamper, R. L. Visual acuity and its measurement. *Ophthalmol. Clin. N. Am.* **16**(2), 155–170 (2003).
24. Levenson, J. H. & Kozarsky, A. Visual acuity change. In *Clinical Methods: The History, Physical, and Laboratory Examinations 3rd Butterworths* (1990).
25. Clare, G., Pitts, J. A., Edgington, K. & Allan, B. D. From beach lifeguard to astronaut: Occupational vision standards and the implications of refractive surgery. *Br. J. Ophthalmol.* **94**(4), 400–405 (2010).
26. Early Treatment Diabetic Retinopathy Study Research Group. Early treatment diabetic retinopathy study design and baseline patient characteristics: ETDRS report number 7. *Ophthalmology* **98**(5), 741–756 (1991).
27. Lesmes, L. A. & Dorr, M. Active learning for visual acuity testing. In *Proc. 2nd International Conference on Applications of Intelligent Systems* (2019).
28. Zhao, Y. *et al.* Psychophysical validation of a novel active learning approach for measuring the visual acuity behavioral function. *Transl. Vis. Sci. Technol.* **10**(1), 1 (2021).
29. Ginsburg, A. P. Contrast sensitivity and functional vision. *Int. Ophthalmol. Clin.* **43**(2), 5–15 (2003).
30. Owsley, C. Contrast sensitivity. *Ophthalmol. Clin. N. Am.* **16**(2), 171–177 (2003).
31. Ginsburg, A. P. A new contrast sensitivity vision test chart. *Am. J. Optom. Physiol. Opt.* **61**(6), 403–407 (1984).
32. Pelli, D. & Robson, J. The design of a new letter chart for measuring contrast sensitivity. *Clin. Vis. Sci.* **2**, 187 (1988).
33. Arditi, A. Improving the design of the letter contrast sensitivity test. *Investig. Ophthalmol. Vis. Sci.* **46**(6), 2225–2229 (2005).
34. Hou, F. *et al.* Using 10AFC to further improve the efficiency of the quick CSF method. *J. Vis.* **15**(9), 2 (2015).
35. Shamir, R. R. *et al.* Comparison of snellen and early treatment diabetic retinopathy study charts using a computer simulation. *Int. J. Ophthalmol.* **9**, 119–123 (2016).
36. Chen, Z. *et al.* Sensitivity and stability of functional vision tests in detecting subtle changes under multiple simulated conditions. *Transl. Vis. Sci. Technol.* **10**(7), 7 (2021).
37. Thurman, S. M. *et al.* Predicting individual contrast sensitivity functions from acuity and letter contrast sensitivity measurements. *J. Vis.* **16**(15), 15 (2016).
38. Hou, F. *et al.* Evaluating the performance of the quick CSF method in detecting contrast sensitivity function changes. *J. Vis.* **16**(6), 18 (2016).
39. Lovie-Kitchin, J. E. Validity and reliability of visual acuity measurements. *Ophthalm. Physiol. Opt.* **8**(4), 363–370 (1988).
40. Dougherty, B. E., Flom, R. E. & Bullimore, M. A. An evaluation of the Mars letter contrast sensitivity test. *Optom. Vis. Sci.* **82**(11), 970–975 (2005).
41. Elliott, D. B., Sanderson, K. & Conkey, A. The reliability of the Pelli–Robson contrast sensitivity chart. *Ophthalm. Physiol. Opt.* **10**(1), 21–24 (1990).
42. Patel, P. J., Chen, F. K., Rubin, G. S. & Tufail, A. Intersession repeatability of contrast sensitivity scores in age-related macular degeneration. *Investig. Ophthalmol. Vis. Sci.* **50**(6), 2621–2625 (2009).
43. Rubin, G. Reliability and sensitivity of clinical contrast sensitivity tests. *Clin. Vis. Sci.* **2**(3), 169–177 (1988).
44. Thayaparan, K., Crossland, M. D. & Rubin, G. S. Clinical assessment of two new contrast sensitivity charts. *Br. J. Ophthalmol.* **91**(6), 749–752 (2007).
45. Arditi, A. & Cagenello, R. On the statistical reliability of letter-chart visual acuity measurements. *Investig. Ophthalmol. Vis. Sci.* **34**(1), 120–129 (1993).
46. Kheterpal, S., Jones, H. S., Auld, R. & Moseley, M. J. Reliability of visual acuity in children with reduced vision. *Ophthalm. Physiol. Opt.* **16**(5), 447–449 (1996).
47. Laidlaw, D. A. H. *et al.* Validation of a computerised logMAR visual acuity measurement system (COMProg): Comparison with ETDRS and the electronic ETDRS testing algorithm in adults and amblyopic children. *Br. J. Ophthalmol.* **92**(2), 241–244 (2008).
48. Snellen, H. *Optotypi ad visum determinandum (letterproeven tot bepaling der gezichtscherpte; probetuchstaben zur bestimmung der sehschaerfe)* (Weyers, 1862).
49. Zhao, Y., Lesmes, L. A., Dorr, M. & Lu, Z.-L. Quantifying uncertainty of the estimated visual acuity behavioral function with hierarchical Bayesian modeling. *Transl. Vis. Sci. Technol.* **10**(12), 18 (2021).
50. Kaiser, P. K. Prospective evaluation of visual acuity assessment: A comparison of snellen versus ETDRS charts in clinical practice (An AOS Thesis). *Trans. Am. Ophthalmol. Soc.* **107**, 311 (2009).
51. Zhao, Y., Lesmes, L. A., Hou, F. & Lu, Z.-L. Hierarchical Bayesian modeling of contrast sensitivity functions in a within-subject design. *J. Vis.* **21**(12), 9 (2021).

52. Rohaly, A. M. & Owsley, C. Modeling the contrast-sensitivity functions of older adults. *JOSA A* **10**(7), 1591–1599 (1993).
53. Watson, A. B. & Ahumada, A. J. A standard model for foveal detection of spatial contrast. *J. Vis.* **5**(9), 6 (2005).
54. Zheng, H. *et al.* Measuring the contrast sensitivity function using the qCSF method with 10 digits. *Transl. Vis. Sci. Technol.* **7**(6), 9 (2018).
55. Haymes, S. A. *et al.* The letter contrast sensitivity test: Clinical evaluation of a new design. *Investig. Ophthalmol. Vis. Sci.* **47**(6), 2739–2745 (2006).

### Acknowledgements

This research was supported by the National Eye Institute (EY017491).

### Author contributions

Z.L.L. designed the study. Y.Z. conducted the simulations. Z.L.L. drafted the manuscript and prepared the figures. All authors reviewed and revised the manuscript.

### Competing interests

LAL and MD have intellectual and equity interests, and hold employment in Adaptive Sensory Technology, Inc. (San Diego, CA). ZLL has intellectual and equity interests in Adaptive Sensory Technology, Inc. (San Diego, CA) and Jiangsu Juehua Medical Technology, LTD (Jiangsu, China). YZ has no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to Z.-L.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023