METHODS & DESIGNS

Sampling distributions of the entropy in visual scanning

ROBIN S. WEISS, ROGER REMINGTON, and STEPHEN R. ELLIS
NASA/Ames Research Center, Moffett Field, California

Identifying true statistical dependencies in visual-scanning data involves showing that the observed scanning pattern is significantly more ordered than that which would be produced by a stratified random-sampling model. In the past, entropy has been used as the index to measure statistical order or dependency. Due to the unknown nature of the underlying sampling distributions of entropy, however, researchers have had to use relatively less powerful nonparametric statistical tests to determine significance. In this paper we present relevant portions of the family of sampling distributions of entropy and show that they are sufficiently normally distributed to allow the use of a more powerful parametric statistical test when attempting to distinguish among the different models of sampling.

We seem to view the world in a continuous fashion, but in actuality we do not. Our eyes fixate features in the visual field, and the distribution of these fixations is usually not uniform (Papin, Naureils, & Santucci, 1980; Stark & Ellis, 1981). Several investigators (e.g., Wewerinke, 1981) have shown that oculomotor information-seeking during instrument monitoring can be modeled as random, or stratified random, sampling. Using similar analyses, other studies (e.g., Kapoula, 1983; Rayner & Pollatsek, 1981) have found structure in the transition patterns of scanning data, which may mean that subsequent fixations are being directed by repetitive, ongoing, cognitive processing.

To identify statistical dependencies in the distribution of fixations among, for example, the instruments in an instrument-monitoring task, three possible patterns of scanning among points of interest to sample visual information need to be considered. Assuming that all sampling is done with replacement, the first case is a random scanning pattern in which there is an equal probability of fixating each point. The second case is a stratified random scanning pattern in which there is a greater probability of fixating certain points. This case can be described as a zero-order Markov process. Senders, Elkind, Grignette, and Smallwood (1966) showed that the probability of a transition between any two distinct points i and j [$p(i \text{ to } j)$], and the conditional probability of a transition to j given a current fixation on i [$p(j|i)$], are calculable from the zero-order, or individual fixation, probabilities of i and j [$p(i)$, $p(j)$]. These equations will be

discussed further in relation to their use by Ellis and Stark (1986).

The third case is a statistically dependent scanning pattern. This occurs when some of the transition probabilities between pairs of points are statistically dependent in such a way that the probability of a transition to any point depends on the current fixation point. These deviations from statistical independence are characteristic of a first-order, or higher Markov, process in which $p(i \text{ to } j)$ and $p(j|i)$ are not calculable from $p(i)$ (Ellis & Stark, 1986). Note that, in the case of stratified random sampling, by chance alone there will be more transitions between the points with greater probabilities. This effect on the scanning pattern may produce an impression of ordered scanning. However, true statistical dependencies do not necessarily exist between those points (Carpenter & Just, 1978). To make such a claim, it must be shown that the observed scanning pattern is more ordered than that which would be produced by a stratified random sampling model.

Ellis and Stark (1986) attempted to demonstrate such a scanning pattern by calculating predicted transition-frequency matrices using adaptations of the equations cited by Senders, Elkind, Grignette, and Smallwood (1966) for describing stratified random sampling. Senders et al. pointed out that, in a scanning task, if the current fixation point is i , there is a probability $p(i)$ that the next fixation point will also be i . Because this is a nonobservable transition, observed transition frequencies between points, $f(i \text{ to } j)$, will be less than those predicted by the commonly used equation

$$\begin{aligned} f(i \text{ to } j) &= N \cdot p(i \text{ to } j) \\ &= N \cdot p(i)p(j), \end{aligned} \quad (1)$$

where N is the total number of transitions among all points

This project was done at the NASA/Ames Research Center under Cooperative Agreement NCC-2-37 to the San Jose State Foundation. Requests for reprints should be sent to Stephen R. Ellis, NASA/Ames Research Center, MS 239-3, Moffett Field, CA 94035.

of interest. Therefore, to calculate a matrix of predicted transition frequencies which would approximate one observed in an actual experiment, transitions from points to themselves must be excluded from the predicted matrix. The denominator of Equation 2 below accomplishes this.

$$p(i \text{ to } j) = \frac{p(i)p(j)}{1 - \sum_{i=1}^f p(i)^2}, i \neq j. \quad (2)$$

By adding the condition $i = j$, the probability space is redefined such that all $p(i \text{ to } j)$ are now conditional probabilities, $p(i \text{ to } j|i=j)$. The N in Equation 1 now represents the total number of transitions among all f points of interest, excluding self-transitions. Using chi-square analyses, Ellis and Stark found that the differences between the observed and predicted transition patterns were statistically significant. However, since a chi-square test is nondirectional, the question of whether this deviation was in the direction of statistically dependent, or purely random, sampling could not be addressed. To make this distinction, Ellis and Stark measured the amounts of statistical dependency (H) in observed transition matrices, and compared them to the amounts in corresponding transition matrices predicted using a stratified random sampling model. The equation

$$H = - \sum_{i=1}^f p(i) \left[\sum_{j=1}^f p(j|i) \log_2 p(j|i) \right], i \neq j \quad (3)$$

was used to calculate this conditional H (Brillouin, 1962), called entropy.

The conditional transition-probability matrices were calculated using the equation

$$p(j|i) = [N/n(i)] \cdot p(i \text{ to } j), \quad (4)$$

where $n(i)$ is the number of transitions from point i . Entropy is directly related to the probability with which a system occupies various states. In a transition matrix, a small H suggests dependencies between the fixation points, whereas a large H suggests a random scanning pattern. Using two-tailed sign tests, Ellis and Stark (1986) found that observed entropies were significantly lower than those predicted by the entropy equation. Thus, the statistically significant deviations found in the chi-square tests were in the direction of statistical dependency.

The sign test used by Ellis and Stark (1986) is a non-parametric, and thus relatively less powerful, statistical test. It was used because of the unknown nature of the underlying sampling distributions of entropy. This paper presents relevant portions of the family of sampling distributions of the H statistic, approximated from Monte Carlo simulations. A more detailed version is presented in Weiss (1987).

METHOD

Design

Twenty-four representative distributions were used to estimate the family of sampling distributions of entropy.

Each of these distributions is described as a population, to emphasize the variable nature of the parameters that define them. The population parameters used in the simulations were selected to correspond to the actual values used or obtained under experimental conditions. One of the variable population parameters was f , the number of fixation points of interest. Fewer than three points will not produce transition matrices adequate for analysis, and previous studies (e.g., Ellis & Stark, 1986) have found that when given an entire display to view, people tend to cluster their fixations around seven or fewer points. Therefore, f took on the values 3, 4, 5, 6, and 7.

One of the constant parameters was N , the total number of transitions between all possible pairs of points in a sample, excluding self-transitions. To ensure stability, N was considered adequate for a population when the predicted transition-frequency matrix had at least 10 transitions in each cell (i.e., $\min [f(i \text{ to } j)] > 10$). The minimum number of transitions was set at 10 because that number surpassed the criteria for a chi-square goodness-of-fit test. The value of N was obtained by first calculating $p(i \text{ to } j)$ using Equation 2. We then set $f(i \text{ to } j)$ equal to 10 in Equation 1 and solved for N . To avoid confounding the results by having unequal sample sizes, N was calculated for each population, and the largest N necessary to satisfy the requirement was used for all populations. As described in the results section, N equaled 858.

The sampling distributions were based on m independent Monte Carlo simulations. Each simulation produced a sample from one of the specified populations. Therefore, we can say that m samples were taken from each population. Monte Carlo simulations generally entail thousands of iterations. However, due to constraints on computer usage and time, we had to choose a moderate value for m while still ensuring stability in the sampling distributions. To establish the minimum value for this constant parameter, we varied only the parameter and ran simulations until there was less than a 5% difference between the mean and standard deviation of the H distribution obtained from the current simulation and the mean and standard deviation of the H distribution obtained from the previous simulation (i.e., simulations were run until the sample statistics of the distributions reached an asymptote). This procedure, detailed in the results section, determined the sample size, m , to be 30. However, sample sizes of 100 were used to ensure further stability in the sampling distributions.

H is appreciatively affected by the zero-order probabilities of the fixation points. If only a few points dominate the fixation distribution, a small H may be calculated but genuine statistical dependencies may not exist. Therefore, $\langle p(i), p(j), \dots p(f) \rangle$, the vector of zero-order probabilities was also a parameter for the simulation. As can be seen in Table 1, the vector varied systematically, from diverse (where one point has a high probability of fixation while the others have low probabilities) to uniform, where all points have approximately equal probabilities of fixation. For eye-scanning data, a probability below .1 is like not having that point of interest, and a probabil-

Table 1
 $\langle p(i), p(j), \dots p(f) \rangle$ Grouped by Number of Fixation Points (f)

Distribution Number	Zero-Order Probability Vector
3 Fixation Points:	
1	$\langle .1 .1 .8 \rangle$
2	$\langle .1 .3 .6 \rangle$
3	$\langle .2 .2 .6 \rangle$
4	$\langle .2 .4 .4 \rangle$
5	$\langle .3 .3 .4 \rangle$
4 Fixation Points:	
6	$\langle .1 .1 .1 .7 \rangle$
7	$\langle .1 .1 .3 .5 \rangle$
8	$\langle .1 .2 .2 .5 \rangle$
9	$\langle .1 .3 .3 .3 \rangle$
10	$\langle .2 .2 .2 .4 \rangle$
5 Fixation Points:	
11	$\langle .1 .1 .1 .1 .6 \rangle$
12	$\langle .1 .1 .1 .2 .5 \rangle$
13	$\langle .1 .1 .1 .3 .4 \rangle$
14	$\langle .1 .1 .2 .2 .4 \rangle$
15	$\langle .1 .1 .2 .3 .3 \rangle$
6 Fixation Points:	
16	$\langle .1 .1 .1 .1 .1 .5 \rangle$
17	$\langle .1 .1 .1 .1 .2 .4 \rangle$
18	$\langle .1 .1 .1 .1 .3 .3 \rangle$
19	$\langle .1 .1 .1 .2 .2 .3 \rangle$
20	$\langle .1 .1 .2 .2 .2 .2 \rangle$
7 Fixation Points:	
21	$\langle .1 .1 .1 .1 .1 .1 .4 \rangle$
22	$\langle .1 .1 .1 .1 .1 .2 .3 \rangle$
23	$\langle .1 .1 .1 .1 .2 .2 .2 \rangle$
24	$\langle .142857 .142857$ $.142857 .142857$ $.142857 .142857$ $.142858 \rangle$

ity greater than .8 for any one point would mean that the probabilities of the other points are negligible. For this reason, the zero-order probabilities assigned to the points within each vector ranged from .1 to .8.

Materials

The Monte Carlo simulations were conducted using a Pascal program with a pseudorandom number (with replacement) algorithm included as a declared function. The algorithm contained a multiplier of 25,173, a modulus of 65,536, and an increment of 13,849 (Cooper & Clancy, 1985). The program was run on the UNIX system with a Pascal compiler.

Procedure

The program was run for each of the populations shown in Table 1. To construct the observed transition matrix for each of the m samples per population, N random numbers between 0 and 1 were generated. These numbers represented transitions in the following way. Possible values of the numbers were divided into f intervals according to the vector of zero-order probabilities. Fixation points were then assigned to each interval so that each random number could be associated with one of the points.

This allowed each number, with the exception of the first, to represent a transition from the point associated with the previous random number to the point associated with the current number. Based on the assumption made by Senders, Elkind, Grignette, and Smallwood (1966) that transitions from all points to themselves occur but cannot be observed, a random number that fell into the same interval as the previous number was discarded.

Next, the observed transition matrices were converted into transition-probability matrices by dividing them by N . These transition-probability matrices were then converted into conditional transition-probability matrices using Equation 4. The H value of each conditional transition-probability matrix was then calculated by Equation 3 and used as a sample statistic.

The sampling distributions of the entropies obtained from the simulations for each population make up a portion of the family of sampling distributions of the H statistic.

RESULTS

As mentioned earlier, the number of samples taken from each population, m , was 100. This choice was based on the results of preliminary simulations in which five fixation points, 1,000 transitions between all possible pairs of points, and a zero-order probability vector of $\langle .1 .1 .3 .4 \rangle$ were arbitrarily selected as the values of the constant parameters. These parameter values could be arbitrarily selected because their purpose at this time was solely to determine the critical value for m , the sample size. The values of m began at 5 and were increased by five for each subsequent simulation. The sample size at which the H distributions stabilized was approximately 30.

We also mentioned earlier that the total number of transitions between all possible pairs of points was determined to be 858. The results of that procedure also showed that the most diverse vector of zero-order probabilities for each f required the largest N , and as the value of f increased so did the value of N .

The first analysis done was a test of the simulation procedure. Simulations were conducted with all parameters held constant except N . Observed transition-probability matrices were constructed using values of N ranging from 10 to 1,000. One sample per population, five fixation points, and a zero-order probability vector of $\langle .1 .1 .1 .3 .4 \rangle$ were arbitrarily selected as the values of the constant parameters. These matrices were then compared to the corresponding matrix predicted using a stratified random sampling model (i.e., Equation 2). We expected that as the number of transitions used to construct the observed matrices (N) increased, the predicted and observed matrices would approach equality. Figure 1 illustrates an example using the cell containing (4 to 1) within these matrices, which shows that this was indeed the case.

One-sample Kolmogorov-Smirnov tests were used to test the normality of the 24 sampling distributions statistically. The H values were standardized and then tested

Figure 1. The probability of a transition over the total number of transitions.

against a standard normal distribution because the means and variances of the population distributions were unknown. Estimation of these population parameters from the samples also meant that the standard tables used for the Kolmogorov-Smirnov test were not valid. The table of critical values given in Dallal and Wilkinson (1986) was used instead to determine the significance levels of the maximum difference (D) statistics obtained from the Kolmogorov-Smirnov tests. The results are presented in Table 2, along with descriptive statistics for each distribution. The hypothesis that the sampling distributions are from a normal population could be rejected in only 3 of the 24 distributions at the .05 level of significance. By chance alone, a few significant deviations would be ex-

pected. In addition, all three of these cases had D values very close to the critical value. It can be seen in Table 2 that, for each distribution, the values of the central tendency measures are approximately equal, and the skews and kurtoses are close to zero. Thus, no significant violations of the assumptions for parametric statistical tests are evident.

An examination of Table 2 reveals some trends in the data. While holding the number of fixation points constant, the means of the sampling distributions tend to increase and the standard deviations tend to decrease as the zero-order probabilities approach equality. However, the vectors with very diverse zero-order probabilities are often the exceptions to these trends. They tend to have the lowest means and the highest standard deviations. The trends are most prominent for the greater number of fixation points. There also appears to be a negative correlation between the number of fixation points and the means. However, the standard deviations are fairly consistent. The skews and the kurtoses of the sampling distributions are also fairly consistent both within and between a number of fixation points. In general, the curves have a slightly negative skew and are slightly platykurtic. An example of a frequency distribution is shown in Figure 2. When taking into account sampling variation, the sampling distributions of H all have the shape of a normal distribution.

DISCUSSION

The obtained sampling distributions for the H index show that the use of a parametric statistical test, such as a t test, would be appropriate for data analysis. This is

Table 2
Descriptive Statistics and Maximum Difference Values (D)
from One-Sample Kolmogorov-Smirnov Tests
for the Sampling Distributions

Distribution Number	Mean	Median	Mode	Skew	Kurtosis	SD	D
1	.899	.901	.896	-.56	-.33	.011	.094*
2	.755	.751	.754	.03	-.13	.026	.070
3	.923	.925	.927	-.12	-.69	.011	.086
4	.932	.932	.938	-.08	-.54	.014	.035
5	.989	.989	.989	-.56	-.21	.005	.110*
6	1.402	1.404	1.406	-.46	-.16	.016	.062
7	1.295	1.295	1.288	-.40	-.22	.030	.057
8	1.418	1.418	1.418	.32	-.12	.020	.045
9	1.456	1.458	1.468	-.07	.26	.019	.075
10	1.528	1.527	1.520	.18	-.21	.013	.084
11	1.769	1.772	1.786	-.22	-.16	.021	.047
12	1.757	1.758	1.742	-.05	-.28	.027	.045
13	1.730	1.730	1.717	-.26	.25	.030	.037
14	1.824	1.826	1.817	-.40	-.02	.024	.069
15	1.835	1.833	1.829	.03	-.48	.024	.073
16	2.084	2.087	2.087	-.74	.81	.025	.064
17	2.107	2.105	2.100	-.08	.23	.024	.053
18	2.105	2.101	2.097	.23	-.04	.026	.074
19	2.174	2.175	2.187	-.26	-.70	.023	.091*
20	2.237	2.237	2.240	-.46	.44	.016	.053
21	2.376	2.375	2.375	-.11	.10	.026	.053
22	2.420	2.420	2.421	-.16	.04	.021	.051
23	2.474	2.472	2.466	-.06	.27	.019	.086
24	2.556	2.557	2.561	-.48	-.35	.007	.083

* $p < .05$.

Figure 2. An example of a distribution from the family of sampling distributions of entropy.

because the assumption of normality would not be seriously violated (Minium & Clarke, 1982). For example, using two-tailed t tests, we reanalyzed the data of Ellis and Stark (1986) and found that observed entropies were significantly lower than those predicted by the entropy equation at the .05 level. These results confirmed those found using the sign-tests. It will now be possible to identify true statistical dependencies in the distributions of fixations in an instrument-monitoring task, or in any other paradigm in which entropy is measured, by distinguishing among the different models of sampling.

Although this paper presents descriptive statistics and significance tests for only a portion of the family of sam-

pling distributions of entropy, more extensive tables are available (Weiss, 1987).

REFERENCES

- BRILLOUIN, L. (1962). *Science and information theory* (2nd ed.). New York: Academic Press.
- CARPENTER, P., & JUST, M. (1978). Eye fixations during mental rotation. In J. W. Senders, D. Fisher, & R. A. Monty (Eds.), *Eye movements and higher psychological functions* (pp. 115-134). Hillsdale, NJ: Erlbaum.
- COOPER, D., & CLANCY, M. (1985). *Oh! Pascal!* (2nd ed.). New York: Norton.
- DALLAL, G. E., & WILKINSON, L. (1986). An analytic approximation to the distribution of Lilliefors's test statistic for normality. *American Statistician*, **40**, 294-296.
- ELLIS, S. R., & STARK, L. (1986). Statistical dependency in visual scanning. *Human Factors*, **28**, 421-438.
- KAPOULA, Z. (1983). The influence of peripheral preprocessing on oculomotor programming. In R. Groner, C. Menz, D. Fisher, & R. A. Monty (Eds.), *Eye movements and psychological functions: International views* (pp. 101-114). Hillsdale, NJ: Erlbaum.
- MINIUM, E. W., & CLARKE, R. B. (1982). *Elements of statistical reasoning*. New York: Wiley.
- PAPIN, J. P., NAUREILS, P., & SANTUCCI, G. (1980). Pickup of visual information by the pilots during a ground control approach in a fighter aircraft simulator. *Aviation, Space, & Environmental Medicine*, **5**, 463-469.
- RAYNER, K., & POLLATSEK, A. (1981). Eye movement control during reading: Evidence for direct control. *Quarterly Journal of Experimental Psychology*, **33A**, 351-373.
- SENDERS, J. W., ELKIND, J. I., GRIGNETTE, M. C., & SMALLWOOD, R. (1966). *An investigation of the visual sampling behavior of human observers* (CR Report No. 434). Moffett Field, CA: NASA Research Center.
- STARK, L., & ELLIS, S. R. (1981). Scanpaths revisited. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements: Cognitive and visual perception* (pp. 193-226). Hillsdale, NJ: Erlbaum.
- WEISS, R. S. (1987). *Sampling distributions of the entropy in visual scanning*. Unpublished master's thesis, California State University, San Jose.
- WEWERINKE, P. H. (1981). A model of the human observer and decision maker. *Proceedings of the 17th Annual Conference on Manual Control*, **81**, 557-570.

(Manuscript received January 20, 1988;
revision accepted for publication January 27, 1989.)