Paper presented at the Symposium "Getting UAV Research Off the Ground" (co-chairs: M.J. Ippel and S.E. Watson) at the 42nd Annual Meeting of the HUMAN FACTORS AND ERGONOMICS SOCIETY, Chicago, October 1998.

UAV CAMERA OPERATOR TASK AS A FINITE-STATE MACHINE

Martin J. Ippel
Galaxy Scientific Corporation
U.S. Air Force Research Laboratory
Brooks AFB (TX)

Stephen E. Watson
U.S. Air Force Research Laboratory
Brooks AFB (TX)

ABSTRACT

In this paper the cybernetic theory of finite-state machines is applied to the problem of S-R compatibility in the UAV Camera Directory Task. It is shown that finite-state machines can be empirically tested by contingency tables in which symbols of an input alphabet (rows) are mapped on symbols of an output alphabet (columns) and each state is represented as a different layer of cross-classification. Information theory provides a natural formalism to quantify the information load of a finite-state machine.

In remote aviation, problems with communication and stimulus-response incongruities often stem from the fact that the different members of the UAV team are looking at different monitors, where each monitor has a unique perspective or reference frame. This paper will focus on the cognitive problem the camera operator faces when required to translate an ego-referenced perspective into that of the UAV perspective, where the effects of input actions are defined in relation to the roll (or longitudinal) axis of the UAV. If the UAV camera operator does not succeed in taking the UAV perspective into account, the behavior of the UAV camera may appear quite unpredictable. For example, let us assume the UAV is flying alongside the *Y*-axis of his computer monitor and the UAV camera is in the same direction as the UAV roll axis. Then, the effect of an input action, say, *forward*, is to move the scanned spot further up alongside the *Y*-axis. However, when the UAV is flying alongside the *X*-axis, say, eastward, the same input action now causes the camera to move along the X-axis. Thus, the same input action corresponds with more than one output action.

In this paper we will describe the task of operating the UAV camera as an abstract machine, a formalism derived from cybernetic theory (Davis, Sigal, & Weyuker, 1994; Denning, Dennis, Qualitz, 1978; Minsky, 1967). For our purpose the class of transducer machines is of particular relevance. A transducer is a machine that translates sequences of symbols chosen from an input alphabet into corresponding sequences of symbols from an output alphabet. In the case of the Camera Directory Task, let the input alphabet consist of four symbols, that is, four arrow keys (i.e., the input alphabet $I = \{\leftarrow, \rightarrow, \uparrow, \downarrow\}$. Verbally, these input actions may be labeled as follows: *left* and *right* (turn the UAV camera), *forward* and *back* (change the angle of descend and thus move the scanned area further away or more nearby the UAV). The output actions that make up a finite output alphabet consist of the set of physical transformations of the UAV camera. Since the perspective of the UAV camera operator is the computer monitor[a], it is convenient to define the UAV camera operations in terms of *X*- and *Y*-coordinates of the areas being scanned. For ease of exposition we will assume a finite set of discrete UAV operations, that is, movements across the *X*- or *Y*-axis (i.e., *x*+, *x*–, *y*+, and *y*–) and rotations in discrete units of 90 degrees clockwise or counterclockwise. These rotations will be denoted as *r*+ and *r*–, respectively. Thus, the entire output alphabet of UAV camera operations is represented by the set $O = \{x+, x–, y+, y–, r+, r–\}$.

The apparently incoherent mapping of input symbols onto output symbols can be understood if we consider the UAV camera as a transducer machine with a finite number of states. The set of states *Q* corresponds to the directions of the UAV camera (i.e., 0, 90, 180, and 270 degrees of angular disparity with the *Y*-axis on the monitor). The behavior of this finite-state machine depends on the functions *g* and *h*. The function *g* is the so-called next-state function and is defined over input alphabet *I* and set of states *Q*. For any $q, q' \in Q$ and $s \in I$, we write $g(s, q) \rightarrow q'$, that is, input s takes state *q* into state *q*′. For each *q*, the finite-state system produces a different function from input alphabet to output alphabet. This is the so-called next-output function:

$$h(s, q) \rightarrow o \qquad (1)$$

that is, given an input symbol *s* and a state *q* an output *o* is produced. As the functions *g* and *h* consist of finite sets of productions, the finite-state machine can be specified completely as a simple table (see Table 1). The four states (i.e., the possible directions of the UAV camera) have been labeled q_1, q_2, q_3, and q_4. Upon starting the machine, one of these states will be initialized as the initial state, usually denoted by

[a] The monitor displays a 2-D map with a north-up alignment, also showing the current position and direction of the UAV.

Paper presented at the Symposium “Getting UAV Research Off the Ground” (co-chairs: M.J. Ippel and S.E. Watson) at the 42nd Annual Meeting of the HUMAN FACTORS AND ERGONOMICS SOCIETY, Chicago, October 1998.

Table 1. Transition function of the finite-state machine modeling the UAV Camera Directory Task

	present state:							
	q_1		q_2		q_3		q_4	
input:	next state	next output	next state	next output	next state	next output	next state	next output
↑	q_1	y+	q_2	x+	q_3	y–	q_4	x–
↓	q_1	y–	q_2	x–	q_3	y+	q_4	x+
←	q_4	r–	q_1	r–	q_2	r–	q_3	r–
→	q_2	r+	q_3	r+	q_4	r+	q_1	r+

Note: The symbols q_1, ..., q_4 denote the states of the finite-state machine. The output symbols *r*+ and *r*– denote clockwise and counterclockwise rotations, respectively. The *x*’s and *y*’s represent changes over the *X*- and *Y*-axis, respectively.

q_I. In the UAV Camera Directory task the initial state is the direction that is in alignment with the UAV roll axis. This finite-state machine accepts four input symbols ←, →, ↑, and ↓. The inputs ↑, and ↓ leave the current state unchanged, while the symbols ← and → cause a change of state. Table 1 specifies a deterministic finite-state machine, which produces for every combination of a state *q* and an input symbol *s*, one and only one output symbol *o*.

It can be argued that the number of states of the UAV camera may, in a sense, be infinite. That is, the camera can be rotated into any directional state of a full circle. However, most likely, any particular camera operator will be able to classify this multitude of rotational states only into a finite (and limited) number of categories. Therefore, a finite-state formalism may be expected to provide an adequate description of the operator task.

The paper discusses how the formalism of finite-state machines can be used to objectively quantify cognitive workload involved in operating the UAV camera.

INFORMATION TRANSMISSION IN THE UAV CAMERA TASK

Any time a camera operator takes action to cause a particular camera operation, he faces an uncertainty of which input action should be chosen. Given the present direction of the UAV camera and the location of the area to be scanned a particular camera operation is expected. Only one of the four arrow keys (i.e., input symbols) can cause this camera operation (i.e., output symbol). The amount of selective work required in performing the Camera Directory Task can be divided into two stages. First, the camera operator has to extract information from the task situation, that is, he has to map the task situation onto a camera operation. Table 2 shows how sixteen different combinations of camera directions and object locations map onto six possible camera actions.

Second, the camera operator has to decode the expected camera action into an input symbol (i.e., a camera operator action) to cause that particular action. The decoding scheme that is appropriate here is the finite-state machine specified in Table 1. Notice a difficulty factor involved in utilizing this finite-state machine as a decoding scheme. Although for every combination of a state and an input symbol, one and only one output symbol can be observed, it is also true that a particular output symbol can be associated with more than one combination of state and input symbol.

A natural formalism to express the amount of selective work that a situation requires is provided by Information Theory (Ash, 1965; Attneave, 1959; Krippendorf, 1986; Shannon & Weaver, 1949). Information theory is concerned with a particular notion of information -- *selective information*. The notion of selective information builds upon the intuitive notion that the difficulty of identifying one particular member of a set is a function of the size of that set.

Let the number of symbols of alphabet *S* be denoted by n_s and let the relative frequency of symbol s_i be p_i ($0 \leq p_i \leq 1$). The uncertainty, or entropy, of a given alphabet is defined as $H(S) = -\sum_i p_i \log_2 p_i$. The entropy of a variable is a measure of the variability for (numerical and) non-numerical variables. In the case of no variability, that is, an information source with a symbol $s \in S$ with $p_s = 1.00$, the entropy value necessarily results in $H(S) = 0$.[b]

Table 2. UAV camera actions given the present camera direction and location of object.

Direction Camera	Location Object/Area			
	n	**e**	**s**	**w**
N	*y+*	*r+*	*y–*	*r–*
E	*r–*	*x+*	*r+*	*x–*
S	*y+*	*r–*	*y+*	*r+*
W	*r+*	*x–*	*r–*	*x+*

Note: the capital letters (N, E, S, W) represent the present directions of the UAV camera, the lower case letters (n, e, s, w) represent the locations of the object to be scanned. The *x*’s and *y*’s represent changes over the *X*- and *Y*-axis, respectively. The output symbols *r*+ and *r*– denote clockwise and counterclockwise rotations, respectively.

Consider the set of task situations as an input alphabet, *S*

[b] If $p_s = 1.00$ then $p_i = 0.00$ for all $i \neq s$. Note that $1 \log_2 1 = 0$ and Information theory has adopted the convention of $0 \log_2 0 = 0$.

= {*Nn*, *Ne*, …, *Ww*}, that is, the set of all possible combinations of camera directions and object locations, and an output alphabet *O* as defined before. Further, the process of extracting information from the task situations is represented by a so-called channel matrix $[a_{ij}]$, $a_{ij} = p\ (o_j \mid s_i)$, $i = 1, \ldots, n_S$, $j = 1, \ldots, n_o$. If *S* is a random variable taking the values $s_1, \ldots, s_M$ with probabilities $p\ (s_1), \ldots, p\ (s_M)$, then the output becomes a random variable. The joint distribution of *S* and *O* is given by $P\{S = s_i, O = o_j\} = p\ (s_i)\ p\ (o_j \mid s_i)$, $i = 1, \ldots, n_s$, $j = 1, \ldots, n_o$. The distribution of *O* is given by $P\{O = o_j\} = \Sigma_i\ p\ (s_1)\ p\ (o_j \mid s_i)$. Thus, the specification of an input distribution leads quite naturally to a joint distribution and an output distribution. Following the definition of entropy, *H* (*S*), *H* (*S* | *O*), and *H* (*O*) can be calculated.

The predictability of the output symbols, $o \in O$, given the set of input symbols, $s \in S$, is defined as *information transmission*. The amount of information transmission may be expressed in several different ways that are mathematically equivalent (Krippendorff, 1986). We adopt the conception that defines information transmission, *T* (*S* | *O*), as the difference between entropy perceived by the operator and that part of this entropy which is noise:

$$T\ (S, O) = H\ (O) - H\ (O \mid S) \qquad (2)$$

where the entropy perceived by the operator is defined as the entropy of the set of output symbols, *H* (*O*), and *H* (*O* | *S*) denotes entropy in *O* given *S*. *H* (*O* | *S*) represents a measure of error, or noise in extracting the required camera action from the task situation. In the case of a perfect extraction process, there is a one-to-one mapping of the input symbols to the output symbols and *H* (*O* | *S*) = 0, in which case *T* (*S*, *O*) = *H* (*O*). In the case of no reliable extraction at all, *H* (*O*) = *H* (*O* | *S*) and *T* (*S*, *O*) = 0.

Throughout this paper it will be assumed that the information extraction requirements in the Camera Directory Task are not beyond the processing capacity of the camera operator. Thus, the probability of error will be arbitrarily small in stage 1, and therefore, it will be assumed that *H* (*O* | *S*) = 0, that is, output symbols $o \in O$ are perfectly predictable from the input symbols $s \in S$. Now, analyzing the outcomes of the decoding stage (i.e., the input symbols chosen by the operator) can assess the degree of control of the UAV camera by the camera operator.

Consider a two-way cross-classification table of expected symbols (i.e., $e \in S_E$) by observed symbols (i.e., $o \in S_O$). The set of expected symbols refers to the symbols *required*, given the camera directions and area locations that the operator encountered during task performance. The set of observed symbols denotes the symbols *actually chosen* by the operator to cause these camera operations. In this two-way cross-classification, *H* (*E*) and *H* (*O*) represent the entropies of the expected symbols and observed symbols, respectively. Notice that both the expected symbols and the observed symbols are elements of the same input alphabet (or, the same output alphabet).

$$T\ (E, O) = H\ (O) - H\ (O \mid E) \qquad (3)$$

where *H* (*O* | *E*) represents a measure of error, or noise, in maneuvering of the UAV camera.

Equation (3) can easily be extended to apply to a model in which a confusion matrix $[b_{jk}]$, with $b_{jk} = p(o_k \mid e_j)$, depends on a state q_i, that is, a function describing the distribution of observed symbols $o \in S_O$ from the distribution of expected symbols $e \in S_E$, given the state $q \in Q$. Now, consider the transmission between *O* and *E* while *Q* is restricted to q_i:

$$T\ (E, O \mid q_i) = H\ (O \mid q_i) - H\ (O \mid E \cap q_i) \qquad (4)$$

The observation design corresponding with Equation (4) is an $I \times J \times K$ cross-classification of states q_i, $i = 1, \ldots, n_i$, expected symbols s_j, $j = 1, \ldots, n_j$, and observed symbols o_k, $k = 1, \ldots, n_k$. For each $q_i \in Q$ a two-way cross-classification of expected symbols and observed symbols can be generated and $T\ (E, O \mid q_i)$ can be computed. The overall conditional information transmission is the weighted average of these two-way statistics over *Q*, that is, $T\ (E, O \mid Q) = \Sigma_i\ P(q_i)\ T\ (E, O \mid q_i)$.

In the sequel of this paper we will present an experiment in which information transmission was analyzed as a function of angular disparity and time limits.

METHOD

Participants

Participants were 173 US Air Force recruits at Lackland AFB ranging in age from 17 to 22 years selected randomly on the sixth day of their Basic Military Training. They participated in the experiment as part of a routine assessment of cognitive skills.

Experimental Task

The recruits participated in a Spatial Maneuvering Test (SMT). In this test a monitor displays a 2-D grid of horizontal and vertical lines, and a triangular icon placed in the center of it. The testee has to use the arrow keys to maneuver the icon toward a flashing light, occurring north, east, south, or west of the icon. The structure of the task was identical to the finite-state machine described in this paper. The participants were randomly assigned to one of nine experimental conditions generated by a factorial design with two factors completely crossed. The first factor will be dubbed MAIN AXIS. This factor refers to the angular disparity between the main axis of the computer monitor and that of the grid. MAIN AXIS had three levels, that is, zero degrees of angular disparity, 30 degrees and 60 degrees, respectively. The icon was rotated in discrete units of ninety degrees (clockwise or counterclockwise). Thus, the angular disparity of the icon with the *Y*-axis was 0, 90, 180 and 270 degrees in the first group, 30, 120, 210 and 300 degrees in the second group, and 60, 150, 240 and 330 degrees in the third group. Per participant the order of presentation of the different angular disparity conditions was randomized. Each subject did 10 trials per angular disparity condition.

The second factor was TIME LIMIT, that is, the time available to accomplish a single trial. This factor has three levels, 2 secs, 4 secs, and 6 secs.

Procedure

Participants were tested in groups of 40 at a time, with each participant at a randomly assigned computerized test station. The SMT was preceded by a practice period of two minutes. Testing time was approximately 10 minutes. After completion of the SMT, participants were administered a test from the USAF Cognitive Ability Measurement battery. Then, a second block of 40 trials was presented to the participants.

Results

To investigate whether the different time limits affected the amount of information transmission (*T*), we calculated *T*, aggregated over individual participants, for each of the twelve conditions of angular disparity within each time limit condition. A repeated measures analysis of variance was performed with MAIN AXIS and TIME LIMIT as between-subjects factor and EXPERIENCE (block 1 versus block 2) and ROTATION (0, 90, 180, and 270 degrees of disparity with the grid's main axis) as within-subject factors. The analysis indicated that EXPERIENCE did significantly effect performance ($F(1, 4) = 38.84$, $p < .01$), but that it did not interact with any of the other experimental factors. TIME LIMIT did not reach a conventional level of significance ($F(2, 4) = .588$, n.s.). However, the power of the latter test was low (p <. 10). Furthermore, the 2- and 4-seconds time limit conditions showed no evidence of a speed-accuracy trade-off. Correlations between accuracy and response times per individual were non-significant, both for the first block of trials and for the second. In the 6-seconds time limit condition these correlations were significant (i.e., r = -.407, and r = .-.364 (df = 38, p < .05) for block 1 and block 2, respectively). This indicated a modest tendency to sacrifice accuracy for speed, or vice versa (between 13% and 16,5% of the performance variance) in the 6-seconds time limit condition.

Based on these analyses we decided to treat the participants in the different time limit conditions as being sampled from one single population and calculated *T* scores aggregated over all participants. To investigate the stability of performance on the SMT, performance in the first and second block of trials was compared. A one-sided paired t-test showed that performance improved with experience (t = -5.462, df = 11, p < .001), but that the pattern of information transmission over the various disparity conditions remained remarkably similar (r = .960, n = 12).

Subsequently, the SMT data were fitted to a series of related nonlinear models of the general form:

$$T = a + b \times \cos(c \times \theta + d) \qquad (5)$$

where θ denotes the radians corresponding with the angular disparity (assessed by clockwise rotation), *a* denotes the intercept, *b* the amplitude, *c* denotes the period of the cosine wave function, and *d* the *X*-axis shift parameter. The models were fitted to the data using the Levenberg-Marquardt estimation method. The loss function to be minimized was the sum of squared residuals. Success of the method depends on the choice of good starting values.

The initial hypothesis assumes that the amount of selective information is constant over the different orientations of the arrow. Information loss, and thus information transmission, can differ per orientation because the processor's processing capacity is partly depleted by the need to mentally rotate (and sustain) a mental representation of the input-output mappings into that direction. This hypothesis predicts information transmission to be a function of the absolute value of the angular disparity between the present direction of the icon and the main axis of the computer monitor. This model (model 1) constrains the values for the period (*c*) and shift parameter (*d*) at 1.00 and 0.00, respectively. In this way, a perfect symmetry is predicted in the decrease of information transmission with a minimum amount of information transmitted at 180 degrees of angular disparity. A X^2 test confirmed what visual inspection of the data already suggested, that is, that the data made this model very unlikely (i.e., $X^2_{(10)} = 10.65$, $p = .385$; parameter values for *a* and *b* were estimated 0.888 and 0.062, respectively).

The next plausible model (model 2a) embodies the hypothesis that a different spatial operation takes over somewhere between 150 and 210 degrees of angular disparity: mental flipping of the icon (or camera orientation). Such a strategy change can take place if a mentally flipping from 0 to 180 degrees requires less workload than mental rotation. This implies that the mapping of output symbols onto input symbols is most difficult when the icon is rotated in a direction that is orthogonal to the computer monitor's main axis (i.e., 90 and 270 degrees of angular disparity). This model constrained the period parameter (*c*) at 2.00. The parameter value of *d* was constrained at 0.00, the parameters a and b were treated as free parameters. This model fitted the data considerably better (i.e., $X^2_{(10)} = 5.34$, $p = 0.861$; $R^2 = 0.506$, while the parameter value for *a* and *b* were estimated 0.888 and 0.247, respectively). A subsequent model (model 2b) differed from model 2a in that the period parameter (*c*) was a free parameter. Freeing the period parameter (*c*) did not really improve model fit (i.e., $X^2_{(9)} = 5.01$, $p = 0.834$; $R^2 = 0.545$). The parameter estimates for *a*, *b*, and *c* were 0.896, 0.257, and 1.918, respectively. We decided to accept model 2a as the best description of the observed data.

Table 3. Comparison of different model fits

	X^2	df	p	ΔX^2	p
model 2a[1]	11.36	12	0.497		
model 2a-a	6.79	11	0.816	4.57	0.03
model 2a-b	10.84	11	0.457	0.52	n.s.

[1] model fit of model 2a on block 2 data

Subsequently, model 2a was fitted to the data of the second block of trials (i.e., all parameter values were fixed at the values of model 2a). Thereafter, we removed the constraint from either the intercept parameter (model 2a-a) or the amplitude parameter (model 2a-b). This analysis showed that the learning effect in the data could be attributed to the intercept of the cosine wave function, rather than to the decreased amplitude (see Table 3).

DISCUSSION

The initial hypothesis that information loss in the UAV camera task would increase as a function of the absolute value of the disparity between the present direction of the camera and the main axis of the computer, could not be confirmed. This suggests that mental rotation is not the mechanism or at least not the only mechanism by which input-output mappings are achieved. Other spatial operations, in addition to rotation, could have been applied to facilitate the mapping of desired output actions to the input actions to cause them. For example, individuals could flip the orientation of the camera. This mental operation would make the difficulty level of the 180 ± 30 degrees angular disparity conditions more similar to the 0 ± 30 degrees angular disparity conditions. Model 2.a, which incorporated this idea provided a much better description of the data. However, the model is most discrepant with the data between 150 and 210 degrees of angular disparity in both trial blocks (see Figures 1 and 2). The data show a monotonous decrease in information transmission between 150 and 210 degrees of angular disparity. It is unclear what causes this phenomenon.

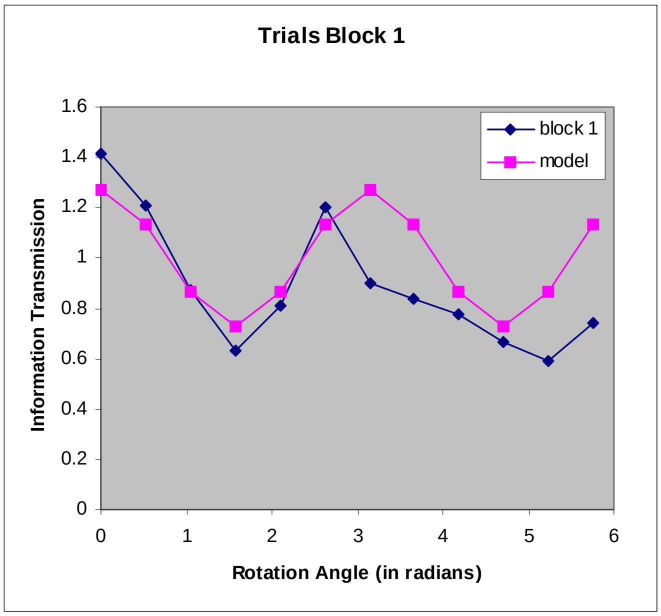

Figure 1. Information transmission as a function of angle of rotation (in radians) in the first block of trials.

References

Ash, Robert, B. (1965). Information Theory. New York: Dover Publications.

Davis, M.D., Sigal, R., & Weyuker, E.J. (1994). Computability, Complexity, and Languages. San Diego, CA: Academic Press.

Denning, P.J., Dennis, J.B., and Qualitz, J.E. (1978). Machines, Languages and Computation. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Ippel, M.J., & Beem, A.L., (1997). Mental Models As Finite-State Machines: Examples and Computational Methods. Armstrong Laboratory Technical Report, AL/HR-TR-1997-0179.

Krippendorff, K. (1986). Information Theory. Structural Models for Qualitative Data. London: Sage Publications.

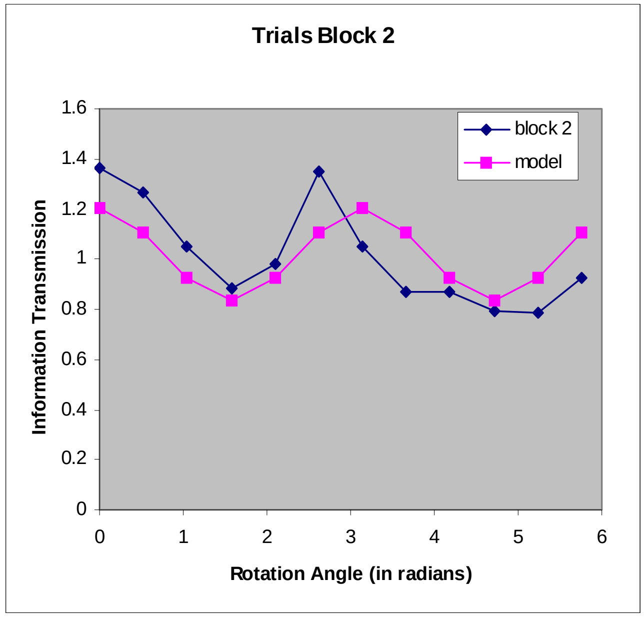

Figure 2. Information transmission as a function of angle of rotation (in radians) in the second block of trials.

Minsky, M.L. (1967). Computation. Finite and Infinite Machines. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Shannon, C.E., & Weaver, W. (1949). The Mathematical Theory of Communication. Urbana: University of Illinois Press.

Paper presented at the 42nd Annual Meeting of the HUMAN FACTORS AND ERGONOMICS SOCIETY
Chicago, October 1998