MODELING INFORMATION OVERLOAD

John E. Boyd and David D. Sworder

ABSTRACT

This paper reports the development of a mathematical model that can be used to simulate information overload. A decision maker is posited, whose task it is to evaluate units of information arriving at random times. The decision maker forms an opinion, based on the (possibly inconsistent) indications of the considered information. This information may or may not be relevant to his or her task, and any unit might or might not accurately indicate the true situation. Determination that a unit of information is irrelevant takes a small amount of time; evaluation of relevant information takes longer. The elapsed time to a prescribed level of confidence provides a metric of decision maker efficiency. The subject's effectiveness depends on how frequently relevant information is presented and to what extent he or she is distracted by the accompanying irrelevant data. The model permits prediction of the optimal rate of presentation of relevant reports as well as analysis of the effects of distracting data on decision maker effectiveness.

John Boyd is a systems scientist with Cubic Defense Systems in San Diego, California. He holds bachelor's and master's degrees in mathematics from California State University at Humboldt and the University of California in Santa Barbara, as well as master's and doctoral degrees in electrical engineering from the University of California in San Diego. His research interests are in nonlinear estimation with application to target tracking, control systems, and hominal modeling.

Dave Sworder is Professor of Electrical and Computer Engineering at the University of California, San Diego. He holds engineering degrees from UC Berkeley and UCLA. Professor Sworder has also served as Dean of Graduate Studies and Research and as Vice Chancellor for Research at UCSD. He is actively involved in studies of novel sensor architectures for control and estimation.

INTRODUCTION

Continuing improvements in sensors, computing power, processing speed, and data dissemination throughput have resulted in vastly increased availability of information to military commanders and decision makers of all types. The trend shows no sign of abating: improved intelligence, surveillance, and reconnaissance (ISR) systems on such platforms as Global Hawk, U-2, Joint STARS, and others are currently proposed, under development, or being deployed. Meanwhile, force reductions and funding constraints combine to limit the number of people available to analyze, interpret, understand, and apply the information presented by ISR systems.

Information Processing

Ideally, data generated by ISR systems should result in knowledge and understanding that will assist the decision maker in carrying out his mission (see Figure 1). Electrical modulation of some sort is typically transformed to data, which can be further reduced to information. Contextualized, interpreted, and absorbed, information becomes knowledge, which, aided by wisdom, leads to understanding. Depending on the sophistication of automatic processing, a human may first enter the processing chain at the data or information block. Unfortunately, no parallel to Moore's law[1] appears to hold for human ability to process, absorb, and store information or knowledge. This fact puts the information processing chain under considerable stress.

The difficulty of absorbing a large amount of data or information is commonly recognized, even proverbial: "like taking a drink out of a fire hose" and "can't see the forest for the trees." After human processing capacity is reached, we tend to ignore the surplus information or, worse, become distracted by it.

Stochastic processes, and in particular, martingale theory, provide a means to model and simulate this phenomenon. This paper describes one approach, based on some earlier work modeling human decision-making. In [SCK93], the authors studied the so-called *order effect* on decision makers: Contrary to intuition, the order in which a set of facts is presented significantly effects the conclusion an evaluator will reach (see also [CS95, CS92, SC94]. This artifact of human behavior, observed in situation evaluators [AB91], was successfully modeled using a recursive estimator not unlike the Kalman filter. The present study, inspired by [SCK93], models the effect of information overload on a situation evaluator, simulating human response to varying amounts of relevant and irrelevant data.

[1]Moore's Law /morz law/ *prov*. The observation that the logic density of silicon integrated circuits has closely followed the curve (bits per square inch)$= 2^{(t-1962)}$) where t is time in years; that is, the amount of information storable on a given amount of silicon has roughly doubled every year since the technology was invented.

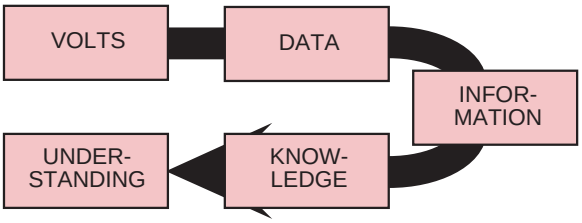

Figure 1: Information processing flow. Sensor readings aid decision maker understanding through a sequence of transformations performed first by machines but ultimately by humans.

MODELING INFORMATION OVERLOAD

A specific example of an information processing problem will make the model development easier to follow. Imagine an image interpreter with a specific area of responsibility (AOR) whose task it is to determine whether a target of interest (TOI) is present in the assigned area. A sequence of images is presented to the interpreter at irregular intervals. Some do not contain the AOR, and after a brief examination, the interpreter discards them. Others contain the AOR, and after careful examination, the interpreter decides whether the relevant image indicates the presence of a TOI. Some will, others may not, perhaps due to screening or countermeasures. The interpreter knows that target indication is imperfect and he or she forms an opinion as to the presence of a TOI as successive images are considered. When a specified level of confidence is reached and confirmed by further images, the interpreter designates the TOI.

The interpreter is trained to ignore any images that arrive during the "dead time" required for preliminary evaluation of image relevance and for analysis and interpretation of those images containing the AOR. While this loss of potentially important information is in some ways regrettable, it is judged preferable to the confusion and inefficiency that might result if analysis were continually interrupted as new images arrive.

One way to model this situation is illustrated in Figure 2. A random number generator produces a

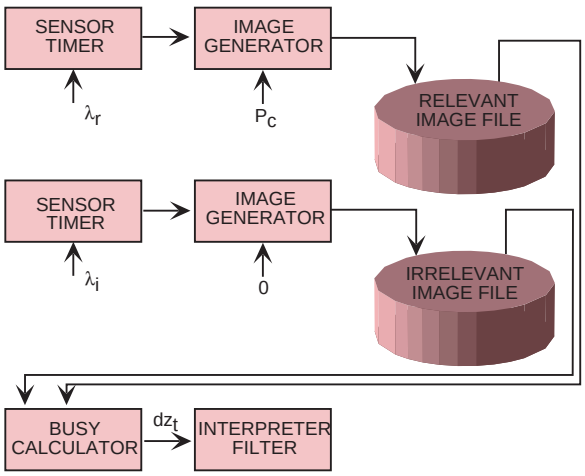

Figure 2: The information overload simulator. It produces relevant and irrelevant data which is merged and decimated to account for interpreter dead time, then passed to a filter model of the interpreter's opinion.

sequence of times at which an image generator is directed at the interpreter's AOR and an image is produced. These images are generated at an average frequency of λ_r frames per second (usually, $\lambda < 1$). The image correctly indicates the presence of a target with probability P_c. A similar system produces irrelevant images (i.e., images not containing the AOR) at a different rate. Since these images will not be evaluated by the interpreter, indication of target presence is not important. In the hypothetical, real situation, the two streams of images would be merged and presented to the interpreter. But to permit us to study the effect of irrelevant information on target detection, we will store both image streams in a file and feed them to the simulated interpreter in the next step.

The image interpreter is modeled by the combination of the last two blocks in Figure 2. The first block, the *busy calculator*, accounts for the dead time resulting from determination of image relevance and analysis of relevant images (T_{dr}). During dead times, any arriving images are discarded. The busy calculator thus forwards to the interpreter only those images he or she is permitted to analyze in accordance with discipline and training. All of them are relevant at this point; those determined not to include the AOR are discarded by the busy calculator. The second block of the interpreter model evaluates each presented image and increases or lowers the interpreter's conviction of target presence as appropriate.

Interpreter Model

Key to this problem is a model for the image interpreter. A bit of mathematics is required to describe it. Let ϕ_t be an indicator vector for the condition *target present*. That is, let $\phi_t \in \{e_1, e_2\}$, the canonical unit two-vectors, $[1\ 0]'$ and $[0\ 1]'$, where $\phi_t = e_1$ signifies that a TOI is present, and $\phi_t = e_2$ the contrary. Assume that ϕ_t satisfies the stochastic differential equation

$$d\phi_t = Q'\phi_t\, dt + dm_t, \tag{1}$$

where Q is a Markov generator matrix and m_t is a purely discontinuous[2] martingale. By modeling ϕ_t as a Markov process, we are assuming it can change state. To model the (in fact, unchanging) state, we will choose a mean state sojourn time that is very long compared to the interpreter's encounter.

The real interpreter examines a sequence of data and forms an opinion about the state of ϕ_t. We can model this as follows. Let the arriving image stream be represented by the process $\{\Delta z_t\}$, where at each time t, $\Delta z_t \in \{[0\ 0]', e_1, e_2, \}$ according to whether no image arrives, an image arrives indicating a TOI is present, or an image arrives indicating no TOI is present. The sum z_t of $\{\Delta z_t\}$ then yields a two-vector whose components represent the number of times a TOI has been indicated (respectively, denied) up to time t. Let the mean rate of arrival of relevant images be given by the parameter λ.

The process $\{z_t\}$ represents a flow of information and gives rise to a mathematical *filtration*,[3] $\{\mathcal{G}_t\}$, that contains all the image-based information. The interpreter's opinion at time t can now be modeled as the expected value of ϕ_t given the information available up to that time, $E[\phi_t|\mathcal{G}_t] \equiv \hat{\phi}_t$.

A recursive scheme to calculate $\hat{\phi}_t$ can be found (see, for example, [Boy96]), but it will require a model of the relationship between the state of ϕ_t and the values of the arriving data. Specifically, we require the *discernibility matrix* D whose ijth component is given by $D_{ij} = P(\Delta z_t = e_i|\phi_t = e_j)$ whenever a nonzero observation Δz_t arrives. This is a model of image fidelity for our problem, and the elements of D represent the probabilities of correct (P_c) or incorrect reports. Specifically,

$$D = \begin{bmatrix} P_c & 1-P_c \\ 1-P_c & P_c \end{bmatrix}.$$

[2]Detailed explanations of these terms can be found, for example, in [Boy96], [SB99], or [Ell82].

[3]Briefly, a filtration $\{\mathcal{G}_t\}$ is an increasing sequence of σ-algebras $\mathcal{F}_t$, each of which may be taken as synonymous for the information available to time t.

Let $R_\phi = \text{diag}(\lambda D \hat{\phi}_t^-)$, and let $P_{\phi\phi} = \text{diag}(\hat{\phi}_t^-) - \hat{\phi}_t^- \hat{\phi}_t^{'-}$. Here and in what follows, the superscript $^-$ or $^+$ indicates the value of $\hat{\phi}_t$ before (respectively after) the observation update. Under these conditions, and with these defined variables, the estimation filter for $\hat{\phi}_t$ is given by the following relationships:
Between observations,

$$\frac{d}{dt}\hat{\phi}_t = Q'\hat{\phi}_t, \tag{2}$$

and at an observation,

$$\Delta\hat{\phi}_t \equiv \hat{\phi}_t^+ - \hat{\phi}_t^- = P_{\phi\phi}\lambda D R_\phi^{-1}\,\Delta z. \tag{3}$$

Between observations, $\hat{\phi}_t$ drifts in the direction $Q'\hat{\phi}_t$. In our problem, such a drift represents a changing opinion as to the presence of a TOI in the absence of data. How quickly this drift would drive $\hat{\phi}_t$ toward its steady-state condition depends on Q. By choosing Q carefully, we will limit the drift.

At an observation, $\hat{\phi}_t$ experiences a step change. This is exactly what we would expect: The arrival of new information confirms or challenges the interpreter's opinion. The actual change is the product of $P_{\phi\phi}$, λD and $R_\phi^{-1}\,\Delta z$. From the definition of R_ϕ, $\lambda D\hat{\phi}_t^-$, we see that R_ϕ represents the mean (expected) vector arrival rate of the data just before the observation arrives. The increment in $\hat{\phi}_t$ therefore depends inversely on the expected data arrival rate: If $\hat{\phi}_t$ is nearly e_1 (the interpreter is almost certain a TOI is present) and a contradictory image $\Delta z_t = e_2$ arrives, the $\hat{\phi}_t$ increment will be relatively large. On the other hand, if $\hat{\phi}_t$ is closer to e_2, the expected arrival rate of images $\Delta z_t = e_2$ will be larger, its reciprocal smaller, and the increment in $\hat{\phi}_t$ will also be smaller as a result. In short, surprising data is given more weight than expected data.

For our case of a two-dimensional probability vector ϕ_t, the third factor, $P_{\phi\phi}$, takes a simple form. If $\phi_t = [p_y\ p_n]$, then

$$P_{\phi\phi} = p_y p_n \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

If the interpreter is uncertain about the presence of a TOI, then $p_y \approx p_n \approx .5$ and the coefficient of $P_{\phi\phi}$ will be about .25. But if he or she is 99% certain in either direction, the coefficient becomes .0099. The effect of $P_{\phi\phi}$ on the update increment of $\hat{\phi}_t$ is to make it larger if the interpreter is uncertain and smaller if the interpreter has already formed a strong opinion. This attribute of the filter mimics real-life "anchoring," the observed difficulty in convincing a person to change his mind, once a commitment has been made.

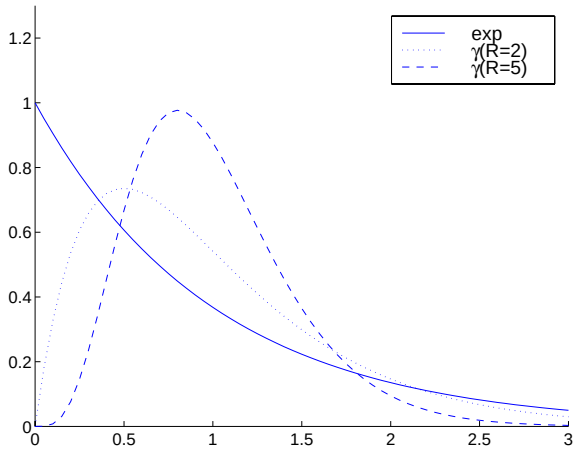

Figure 3: Density functions for an exponential distribution and two members of the gamma family. Although each has mean 1, the gamma-distributed variables are much less likely to be close to 0.

Busy Calculator

We turn now to the issue of dead-time in our model. Let T_i be the average interval in seconds between arriving irrelevant images, and let T_r be the interval for relevant images. Assume the time for determination of image relevance is exponentially distributed (see Figure 3) with mean T_{di} and that the time for analysis and evaluation of relevant images has a gamma (or Erlang) distribution with mean T_{dr}. Observe from Figure 3 that the exponential distribution has a preponderance of short times. We judge this characteristic to be inappropriate as a model of analysis time, choosing instead the gamma model[4] with parameter 4. Although the gamma density family is an extension of the exponential, the other gamma distributions lack the strong tendency toward small values, as is clear from the figure.

The busy calculator draws images from both relevant and irrelevant images streams. When a relevant image is encountered, it is copied to the interpreter input file and the calculator draws a gamma-distributed random time of the prescribed mean, T_{dr}. The clock advances, skipping over both relevant and irrelevant images, and processing re-

[4] See, for example, [Pap91, chapter 4] for explicit density formulas and a discussion of the characteristics of these distributions.

sumes at the new time. A similar skip (of mean length T_{di}) is performed when an irrelevant image is encountered. The resulting image file has thus accounted for classification and analysis dead time.

In the real-world example, the results of an image analysis would first be available to update interpreter opinion at the end of the analysis dead time. For simplicity, we omit this detail from our simulation, allowing the interpreter opinion to drift during the dead time. Because of the very long time constants we choose for the target presence state ϕ_t, however, this effect is negligible.

Alternative Models

The model we have chosen is perhaps the simplest that will fit the problem. Others might be considered. For example, suppose we model the image relevance as random and only indirectly known. Let $\alpha_t \in \{e_1, e_2\}$ be in indicator vector for the condition "radar is pointed at the AOR" so that $\alpha_t = e_1$ indicates the radar is currently delivering relevant images and $\alpha_t = e_2$ indicates the opposite. Let ρ_t be defined as was ϕ_t in our earlier discussion, indicating the condition "TOI is present." Now construct a composite state vector $\phi_t = \alpha_t \otimes \rho_t$, where $\otimes$ indicates the Kronecker product. The values of ϕ_t are shown in Table 1.

The observation sequence $\{\Delta z_t\}$ now takes values in the same space as $\{\phi_t\}$, and a new discernibility matrix D must be found. The Markov generator Q must now permit occasional transitions between relevance and irrelevance. The filter for this problem is given by equations (2) and (3) with the new matrix parameters. But $\hat{\phi}_t$ now reflects the estimated current image relevance as well as the presence of a TOI.

We could also build the busy calculator into the interpreter model where, it could be argued, it belongs. In addition to the $\{\alpha_t\}$ and $\{\rho_t\}$ processes just defined, let $\{r_t\}$ be a gamma process of order, say, R which determines the dead time for the interpreter. The process $\{r_t\}$ takes values in $\{e_1, e_2, \ldots, e_R\}$. State e_1 is the "idle" state: While $r_t = e_1$, the interpreter is waiting for an image to arrive. When an apparently irrelevant image arrives, r_t makes the transition $e_1 \mapsto e_2$. If a relevant image arrives, r_t makes the transition $e_1 \mapsto e_{R+1}$. When not in state e_1, *no arriving images are considered*. It is this rule that builds the dead time into the interpreter model. When not in state e_1, r_t transitions only to the next lower state. Let T_i be the mean sojourn time in state $i, i = 1, 2$, and let λ be the mean arrival rate of the merged stream of images. Suppose $1/3$ of the images are irrelevant and $T_{di} = T_{dr/3}$. Then for $R = 3$ the generator matrix for $\{r_t\}$ is given by

$$Q = \begin{bmatrix} -\lambda & \lambda/3 & 0 & 2\lambda/3 \\ 1/T_{di} & -1/T_{di} & 0 & 0 \\ 0 & 1/T_{di} & -1/T_{di} & 0 \\ 0 & 0 & 1/T_{di} & -1/T_{di} \end{bmatrix}.$$

The derivation of equations (2) and (3) assumes that state transitions in ϕ_t are not coincident with arrival of observations Δz_t. In this model, that may not be the case: Arrival of a new images may be very likely to coincide with changes in ϕ_t. A new derivation is required, and the results yield a somewhat different form for the $\hat{\phi}_t$ filter. The modified filter (and its development) are shown in the Appendix, but it will not be discussed further here.

Table 1: State vector values for an alternative interpreter model incorporating image relevance into the estimated state.

ϕ_t	α_t	ρ_t	*Meaning*
e_1	e_1	e_1	target present, AOR in image
e_2	e_1	e_2	target present, AOR not in image
e_3	e_2	e_1	no target, AOR in image
e_4	e_2	e_2	no target, AOR not in image

EXAMPLE

We implemented the information overload simulator in MATLAB. We considered two cases for this report:

irrelevant data Given a fixed amount of relevant data, how does the interpreter respond to increasingly frequent interruptions by images not containing the AOR?

relevant data Is it possible to have too much relevant data? How much information is gleaned from a sequence of images, all containing the AOR, arriving so rapidly they cannot all be analyzed?

Study 1: Choking on Irrelevant Data

The objective of this study was to determine the effect on TOI detection time of varying amounts of irrelevant data. We set the simulation parameters as shown in Table 2. For a metric we chose time to detection, defined as the first time at which

Table 2: Simulation parameters for Study 1: Choking on Irrelevant Data.

Param.	Value	Remarks
λ_r	$.0167$ /sec	mean rate, relevant data
T_{dr}	30 s	mean dead time, relevant
T_{di}	3 s	mean dead time, irrelevant
P_c	.75	image accuracy
T_d	100 s	time for detection

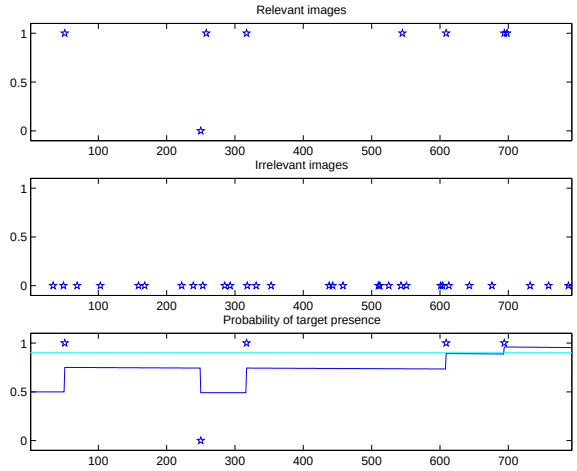

Figure 4: A single run of the information overload simulator. The top graph shows the sequence of relevant observations and their values; the middle graph depicts a similar sequence of images not containing the AOR, and the bottom graph those images that survive the busy calculator. The probability of *TOI present* is also plotted.

the probability of TOI had been the detection confidence threshold, 0.90, for T_d seconds. The mean frequency of arrival (frames per second) of the nonAOR-applicable data was varied from .02 to .5; while frequency of arrival of relevant data was held constant at mean .0167 s. Thus, on average, there was a relevant frame every 60 seconds, and irrelevant frames at a rate of every 2 seconds to every 50 seconds. We performed 200 runs for each value of T_i.

Figure 4 shows the result of a single simulation run similar to those for Study 1. The top graph shows that in this case, 8 relevant images were received, 1 of which erroneously indicated that no TOI was present. The middle graph indicates irrelevant data was somewhat more frequent. The interpreter opinion filter saw, in this case, most of the relevant data, losing 3 of the observations. At

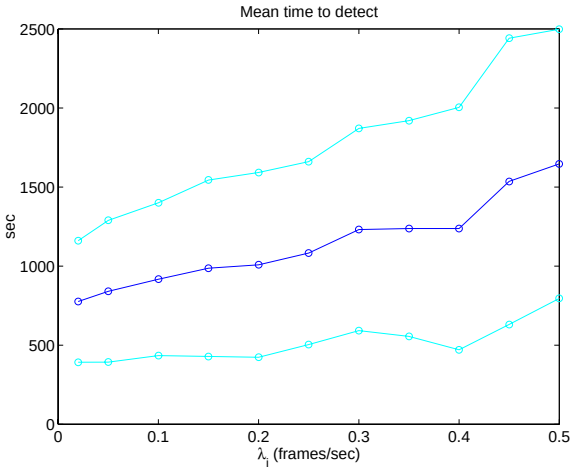

Figure 5: Mean time to detect a TOI for varying frequencies of irrelevant data. The effect of information overload is clear: MTD doubles as the rate of distracting data increases.

about 700 seconds, the interpreter opinion crossed the 90% confidence level, and after remaining there for $T_d = 100$ seconds, a TOI detection was scored at about 800 seconds.

Shown in Figure 5 are the detection times for several rates of arrival of irrelevant information, together with one-standard deviation points about the means. At very low rates of irrelevant data, detection time averaged about 800 seconds. But as the distractions became increasingly frequent, more and more of the relevant data was blocked and the detection time rose to about 1500 seconds. Large variability was evident in the individual runs. This is evident in the relatively large standard deviations and the lack of smoothness in the 200-run means.

Figure 6 displays another way to view the effect of information overload for Study 1. It shows the percentage of relevant images generated that actually reached the opinion filter, the data that survived the busy calculator. As λ_i increases, more of the relevant frames are excluded: At $\lambda_i = .5$ seconds, only about a fourth of them survive: The irrelevant data so overloads the interpreter that he sees only 25% of the good data.

Study 2: Too Much Relevant Data

In the second study, we examine the effect of varying amounts of AOR-relevant data. To simulate this, set the frequency of irrelevant data very low, and vary the frequency of relevant data. We expected that as the interpreter began to be over-

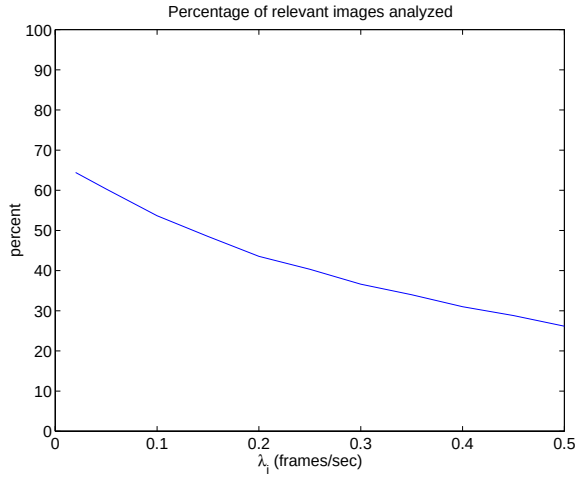

Figure 6: Percentage of relevant images analyzed as a function of T_i for Study 1. As the operator is increasingly overloaded with irrelevant data, fewer of the relevant images can be analyzed.

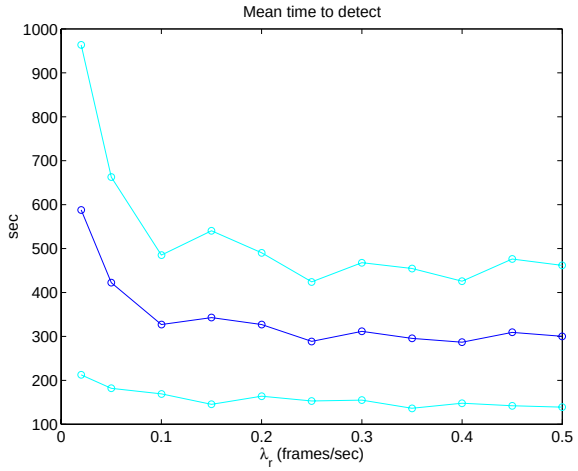

Figure 7: Mean time to detect for Study 2, together with one-standard deviation bounds at each simulated data frequency.

Table 3: Simulation parameters for Study 2: Too Much Relevant Data.

Param.	Value	Remarks
λ_i	$.0001$ /sec	mean rate, irrelevant data
T_{dr}	30 s	mean dead time, relevant
T_{di}	3 s	mean dead time, irrelevant
P_c	.70	image accuracy
T_d	100 s	time for detection

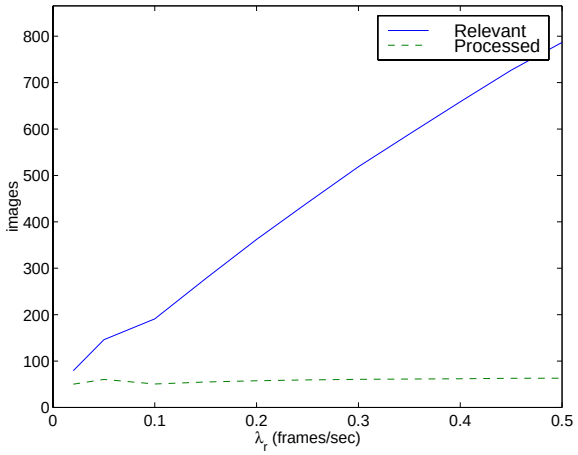

Figure 8: Relevant images generated and relevant images processed, as functions of T_r for Study 2.

loaded with data, even good data, she would have to ignore good data. The simulation parameters are shown in Table 3; they are similar to those used in the first study. The detection threshold remained at 90 percent, and we ran 200 trials at each λ_r.

Figure 7 shows the results of the simulation. For high frequency of relevant data, mean detection time continues to decrease slightly with increasing λ_r. Since our interpreter is perfectly disciplined, she ignores arriving data until she finishes analyzing the image before her. Her effectiveness does not drop. But it is clear that rates above about 0.1 frames/second buy only a limited improvement in MTD.

Shown in Figure 8 is a picture of what is happening to the data. The solid curve represents the number of valid images being generated before detection is scored, while the dashed curve is the number of images examined. The examined image count is nearly constant – what varies is how long it takes for the interpreter to see the requisite number of images. At high data rates, most of the data is ignored by our interpreter. One envisions a workspace littered with discarded, unexamined images. This "information inefficiency" is further illustrated in Figure 9, where we have plotted the percentage of relevant images generated that actually get analyzed. When the data frequency is low, about two thirds of the images make it to the analysis table. But as frequency increases, image attrition exceeds 90%.

This study illustrates the kind of tradeoffs that must be made in designing a real system. If the information source is an airborne sensor, and if the analysis facility is on the ground, then a data link must be used to transmit the sensor outputs. High data capacity in an RF data link implies large anten-

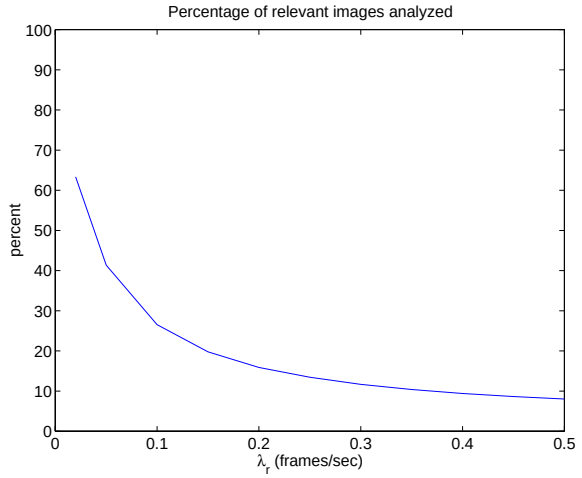

Figure 9: The percentage of relevant images actually reaching the image interpreter for Study 2.

nas, high transmit power, diminished mobility of the ground station, and higher costs. These disadvantages may have to be balanced against the limited improvement in TOI detection time that a full capacity data channel might offer.

CONCLUSION

The martingale calculus-based model we have described appears to have good potential for simulation of human decision making activities and, in particular, the phenomenon of information overload. To enhance its usefulness in a specific, real-world decision-making problem, a calibration against actual human performance is probably required. The model could also provide a useful component in a more complex system simulation.

Several additional cases for analysis suggest themselves. For example, suppose discipline is imperfect and an interpreter abandons an image in progress when a new piece of information appears. It would seem such an interpreter might be completely stymied by too-frequent delivery of information, no matter how valid. Or suppose information is less frequent, and interpreter conviction softens in the occasional long wait for confirming or disconfirming data. These and other problems will be studied as time permits. The authors welcome e-mail correspondence related to this work.

APPENDIX

In this appendix, the state estimation algorithm will be developed.[5] Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and let $\{\mathcal{F}_t\}$ be a right-continuous filtration which generates $\mathcal{F}$. There are several exogenous processes defined on this space, all right-continuous and $\mathcal{F}_t$-adapted: $\{\alpha_t\}$ is the AOR indicator vector and is of dimension K; $\{\rho_t\}$ is the TOI indicator vector and is of dimension L; and $\{r_t\}$ is the observation timing vector and is of dimension R. The comprehensive state is expressible in terms of the primitive processes: $\phi_t = r_t \otimes \rho_t \otimes \alpha_t$. From the state the primitives can be deduced. Let: $F_r = \mathbf{I}_R \otimes \mathbf{1}'_{LK}$; $F_\rho = \mathbf{1}'_R \otimes \mathbf{I}_L \otimes \mathbf{1}'_K$; $F_\alpha = \mathbf{1}'_{RL} \otimes \mathbf{I}_K$; and $F_{\rho\alpha} = \mathbf{1}'_R \otimes \mathbf{I}_{LK}$. Then: $r_t = F_r\phi_t$; $\rho_t = F_\rho\phi_t$; $\alpha_t = F_\alpha\phi_t$; and $\rho_t \otimes \alpha_t = F_{\rho\alpha}\phi_t$. If a particular component of $\{r_t\}$ is desired, it is easily found.

An observation is generated when there is a transition in $\{r_1\}$: If $\{r_t\}$ is such that $e_1 \mapsto e_2$, information blocking is short and if $\{r_t\}$ is such that $e_1 \mapsto e_{R+1}$, information blocking is long. There is a discernibility matrix, D, of dimension $LK \times LK$ which measures the quality of the observation. When $\Delta z_t \neq 0$ the probability that $\Delta z_t = e_i$ is the ith component of the vector $\boldsymbol{\lambda}_t = D(\rho_t \otimes \alpha_t)$. The σ-field generated by $e_i'\phi_t \vee \mathcal{F}_t$ is labeled $\mathcal{F}_t^i$. The observation $\{z_t\}$ generates $\{\mathcal{G}_t\}$ with $\mathcal{G}_t$-innovations process $\{\nu_t\}$. We seek a causal map from $\{\nu_t\}$ to $\{\hat{\phi}_t\}$. This is a difficult construction because of the nonlinear and discontinuous system dynamics.

The modal observations are distributed across the LK topical bins. An observation is received whenever $\{r_1\}$ changes. When $\{r_1\}$ has a jump, $(\Delta r_1)^2 = 1$. Hence, $E[\Delta z_t|\mathcal{F}_t] = E[(\Delta r_1)^2 D R_{\rho\alpha}\phi_t|\mathcal{F}_t]$. The modal process is a purely discontinuous semimartingale with dynamic model:

$$d\phi = Q'\phi\, dt + dm_t$$

Hence $(\Delta r_1)^2 = e_1' F_r \Delta m_t' \Delta m_t F_r' e_1$. But

$$E[\, dm\, dm'|\mathcal{F}_t] = d\langle m, m; \mathcal{F}_t\rangle_t = \boldsymbol{V}(\phi_t)\, dt$$

where $\boldsymbol{V}(\phi_t) = [\text{diag}(Q'\phi_t) - \text{diag}(\phi_t)Q - Q'\text{diag}(\phi_t)]$. [SB99]. The $\mathcal{G}_t$-predictable quadratic variation of $\{m_t\}$ is the expected value of $d\langle m, m; \mathcal{F}_t\rangle_t$: $d\langle m, m; \mathcal{G}_t\rangle_t = \boldsymbol{V}(\hat{\phi}_t)\, dt$

To develop the PME note that $E[dz|\mathcal{F}_t^i] = (e_1 R_r \mathbf{V}(e_i) R_r' e_1) D R_{\rho\alpha} e_i$. Define a matrix h' with ith column $h'_{\cdot i} = (e_1 R_r \mathbf{V}(e_i) R_r' e_1) D R_{\rho\alpha} e_i$. Then

$$dz_t = h'\phi_t\, dt + dn_t \tag{4}$$

[5] For notational convenience in what follows $\boldsymbol{S}$ will designate an integer index set $\{1, \ldots, S\}$, "$*$" is the Hadamard product ($(x * y)_i = x_i y_i$), and the integrands in stochastic differential equations will be understood to be predictable versions of the associated right continuous, random processes.

where $\{n_t\}$ is an $\mathcal{F}_t$-martingale. Similarly

$$dz_t = h'\hat{\phi}_t\,dt + d\nu_t \tag{5}$$

where the innovations process, $\{\nu_t\}$, is a $\mathcal{G}_t$-martingale. The innovations process can be written:

$$d\nu_t = h'\tilde{\phi}_t\,dt + dn_t \tag{6}$$

The processes, $\{n_t\}$ and $\{\nu_t\}$, are purely discontinuous martingales. Both have the same predictable quadratic variation; e.g.,

$$\begin{aligned} d\langle n_t, n_t; \mathcal{F}_t\rangle_t &= E[\Delta z_t\,\Delta z_t'|\mathcal{F}_t] = \mathrm{diag}(\boldsymbol{\lambda}_t)\,dt \\ d\langle n_t, n_t; \mathcal{G}_t\rangle &= \mathrm{diag}(\hat{\boldsymbol{\lambda}}_t)\,dt = R_\phi\,dt, \end{aligned} \tag{7}$$

where (7) is taken to be the definition of R_ϕ. The system will be assumed to be such that each component of $\hat{\boldsymbol{\lambda}}_t$ is positive, and consequently, $R_\phi > 0$. Also $d\langle n_t, \nu_t; \mathcal{F}_t\rangle_t = \mathrm{diag}(\boldsymbol{\lambda}_t)\,dt$ and $d\langle n_t, \nu_t; \mathcal{G}_t\rangle = R_\phi\,dt$. Let $\hat{\boldsymbol{\lambda}}_t^{-1}$ be understood componentwise, and denote $\sum_j (e_j - e_i)Q_{ij}h_{i\cdot}$ by $R^i_{\phi z}$. The PME is given by:

Modal Estimation

Between observations,

$$\frac{d\hat{\phi}_t}{dt} = Q'\hat{\phi}_t - \sum R^i_{\phi z}\hat{\phi}_i\mathbf{1} \tag{8}$$

At a modal measurement,

$$\Delta\hat{\phi} = (P_{\phi\phi}h + \sum_i R^i_{\phi z}\hat{\phi}_i)(\hat{\boldsymbol{\lambda}}^{-1} * \Delta z) \tag{9}$$

Discussion

Decompose the semimartingale $\{\hat{\phi}_t\}$ (see [Ell82, Theorem 18.11]),

$$d\hat{\phi} = E[d\phi_t|\mathcal{G}_t] + \gamma_{\phi\phi}\,d\nu_t \tag{10}$$

where $\gamma_{\phi\phi}$ is a $\mathcal{G}_t$-predictable matrix process. For notational convenience, let $Q'\phi_t = F_\phi$. Then $E[d\phi_t|\mathcal{G}_t] = \hat{F}_\phi\,dt$. So we have

$$d\hat{\phi}_t = \hat{F}_\phi\,dt + \gamma_{\phi\phi}\,d\nu_t \tag{11}$$

To find $\gamma_{\phi\phi}$ explicitly, the formalism used successfully by Elliott in [Ell82] will be employed. Note that

$$E[d(\phi z')|\mathcal{G}_t] = E[d\phi\; z' + \phi\,dz' + d\phi\,dz'|\mathcal{G}_t]. \tag{12}$$

But $E[d\phi\; z'|\mathcal{G}_t] = \hat{F}_\phi z'\,dt$. The second term in (11) can be written

$$E[\phi\,dz'|\mathcal{G}_t] = E[\phi\phi' h\,dt + \phi\,dn'|\mathcal{G}_t]$$

Finally $d\phi_t dz_t' = \Delta\phi_t\Delta z_t'$. But $d\langle\phi_t, z_t; \mathcal{F}_t^i\rangle_t = E[\Delta m_t\phi_t' R'_{\rho\alpha}D'(\Delta r_1)^2|\mathcal{F}_t^i]$. This can be written:

$$d\langle\phi_t, z_t; \mathcal{F}_t^i\rangle_t = \sum_j (e_j - e_i)Q_{ij}h_{i\cdot}\,dt$$

Then $d\langle\phi_t, z_t; \mathcal{G}_t\rangle_t/dt = \sum_i R^i_{\phi z}\hat{\phi}_i$.

Combining these equations, it follows that

$$E[d(\phi z')|\mathcal{G}_t]/dt = R_{\phi\phi}h + \hat{F}_\phi z' + \sum_i R^i_{\phi z}\hat{\phi}_i. \tag{13}$$

We can express $E[d(\phi z')|\mathcal{G}_t]$ in another way. From (10), we have

$$d\hat{\phi}_t\,dz_t' = (\gamma_{\phi\phi}\,d\nu_t)\,dn_t'.$$

It is a direct calculation to show that

$$\begin{aligned} d\hat{\phi}\,z' &= \hat{F}_\phi z'\,dt + d\mu \\ \hat{\phi}\,dz' &= \hat{\phi}\phi' h\,dt + d\mu. \end{aligned}$$

where $\{\mu_t\}$ is a martingale. Collecting the terms,

$$d(\hat{\phi}z') = (\gamma_{\phi\phi}\,d\nu\,dn' + \hat{\phi}\phi' h + \hat{F}_\phi z')\,dt + d\mu,$$

and taking the $\mathcal{G}_t$-expectation of this,

$$E[d(\hat{\phi}z')|\mathcal{G}_t]/dt = \gamma_{\phi\phi}R_\phi + \hat{\phi}\hat{\phi}'h + \hat{F}_\phi z'. \tag{14}$$

The predictable compensators, (13) and (14), must be equal. From this it follows that

$$\gamma_{\phi\phi} = (P_{\phi\phi}h + \sum_i R^i_{\phi z}\hat{\phi}_i)R_\phi^{-1}. \tag{15}$$

Substituting this into (10) yields

$$d\hat{\phi}_t = Q'\hat{\phi}_t\,dt + (P_{\phi\phi}h + \sum_i R^i_{\phi z}\hat{\phi}_i)R_\phi^{-1}\,d\nu_t, \tag{16}$$

Consider a time when there is no observation: $d\nu_t = -\hat{\boldsymbol{\lambda}}_t\,dt$. Then

$$R_\phi^{-1}d\nu_t = -\mathrm{diag}(\hat{\boldsymbol{\lambda}}_t^{-1})\hat{\boldsymbol{\lambda}}_t, dt = -\mathbf{1}\,dt.$$

So

$$hR_\phi^{-1}h'\hat{\phi} = \lambda D'\mathbf{1} = \lambda\mathbf{1}.$$

Also $P_{\phi_i\phi}\mathbf{1} = 0$ for every $i \in \boldsymbol{S}$. Hence, $P_{\phi\phi}hR_\phi^{-1}\,d\nu_t = 0$. The result in (8) and (9) then follows.

Conclusion

For problems in which the modal state is hidden in the modal measurement, a symmetric modification of the PME improves the speed of modal identification. The base-state estimate is less sensitive to modal identification than is often supposed. Only when the application requires high quality modal estimation, is the symmetric PME required.

References

[AB91] L. Adelman and T. Bresnick, *Examining the effect of information sequence on expert judgement*, Proceedings of the Symposium on Command and Control Research Command and Control Research (1991), 306–317.

[Boy96] J.E. Boyd, *Nonlinear filtering for multimodal systems with mode-dependent obsrvations*, Ph.D. dissertation, University of California, June 1996.

[CS92] G.A. Clapp and D.D. Sworder, *Command, control and communications: The human role in military C3 systems*, Control and Dynamic Systems, Advanced Concepts in Tracking (1992), 513–541.

[CS95] G.A. Clapp and D.D. Sworder, *Modeling TAD/TMD decisonmaker dynamic response*, Proc. of the Fire Control Symposium (1995), 529–547.

[Ell82] R.J. Elliott, *Stochastic calculus and applications*, Springer-Verlag, 1982.

[Pap91] A. Papoulis, *Probability, random variables, and stochastic processes*, 3rd ed., McGraw-Hill, Inc, New York, 1991.

[SB99] D.D. Sworder and J.E. Boyd, *Estimation problems in hybrid systems*, Cambridge University Press, 1999.

[SC94] D.D. Sworder and G.A. Clapp, *Decisionmaker styles in dynamic situation assessment*, Proceedings of the Symposium on Command and Control Research Command and Control Research (1994), 591–601.

[SCK93] D.D. Sworder, G.A. Clapp, and T.W. Kidd, *Model-based prediction of decisionmaker delays and uncertainty*, Cybernetics and Systems; An International Journal **24** (1993), 51–68.