Self-management of criticism in dialog:
Dynamic regulation through automatic mediation

Effective dialog requires responsible criticism to insure that deliberatory procedures are followed and more generally to enable the rapid correction of errors so that time can be spent constructively. Self-management of criticism implies both that people offer criticism when appropriate and that they are as a group able to judge the relative competence of their members. An optimal level of criticism permits people to judge each other's relative competence. An automatic mediator which dynamically regulates the level of criticism in dialog, based upon a signal detection model of individual competence and using Bayesian estimation is presented. The mediator receives anonymous offers of criticism from individuals when statements are judged erroneous and combines these signals using model parameters to predict the chances that a given criticism will be sustained by a group consensus. The accuracy of these predictions controls the relative level of reward given to successful criticisms. Inadequate reward tends toward a lack of criticism, "group-think", while excessive reward tends toward every statement being disputed, bitter arguement. Dynamic regulation of rewards is used to guide the group between these extremes, thus facilitating accurate judgement of relative competence and thereby proper treatment of statements in the dialog process.

David Stodolsky, PhD
Institute for Organization and Management
Aarhus School of Business Administration
Ryhavevej 8, DK-8210 Aarhus V, Denmark
(+45 6) 15 96 33, local 43
2-December-1984

Self-management of criticism in dialog:
Dynamic Regulation Through Automatic Mediation.

The purpose of this paper is to illustrate how some old (even
ancient) and well established ideas about how people com-
municate and solve problems can be integrated within the
framework of the new information technology. A system for
self-management which is intended to maximize individual
control and participation in dialog is described. A key element
is a computer-based mediation mechanism which permits full dis-
tribution of control and participation through the collection
of anonymous individual responses and automatic resolution of
conflicting requests to speak.

Such a mechanism,which eliminates the need for a fixed leader-
ship role and provides for a protected mode of expression
(Stodolsky, 1981), greatly simplifies the analysis of,and
support of,group interaction processes. There are many related
regulatory processes which proceed simultaneously in group inter-
action. Some of these have a basically physiological nature and
others are concerned with maximizing decision performance.
Intermediate social and psychological phenomenon integrate
these processes.   This outline will emphasize mutual regul-
ation across this range of processes and among group members.
Specifically, the focus of analysis will be upon the level of
criticism in the dialog process.

The objective is to illustrate how by adjusting this level for
peak group performance, one may simultaneously stabilize social,
psychological, and physiological regulatory processes.

Philosophical foundations.

In evaluating group performance we accept a rational actor theory
of group decision making. As used in an ethical theory, rational
is defined reformatively as "actions, desires, or moral systems

2.

which survive maximal criticism and correction by facts and
logic (Brandt, 1979)." Two components are necessary to achieve
rational action, 1) fully and correctly informed desires, which are
integrated with 2) rational choice mechanisms. Rational choice
mechanisms use all available information and the rules of logic,
both deductive and inductive. It is assumed here that if people
can transmit information about their desires and preferences,
and their bases, freely to each other then correction of irrational
desires will occur. This follows from both the "dissonant" nature
of irrational desires (both intra- and inter-personally) and their
wasteful and costly nature (Brandt, 1979). Rational actions then
follow, given all available information and the application of
the rules of logic. The system we are to describe has powerful
mechanisms to balance participation and thereby make available
all information which members bring to the group. It also contains
a mechanism for promoting criticism, meant to enforce the rules
of deductive logic. Other mechanisms are available to support
inductive logic, but these are beyond the scope of this paper
(Stodolsky, 198o).

Brandt (1979) suggests a moral system based upon a rule utilit-
arian ethical theory which is ideally monistic (based on a
single principle). In practice such a system makes excessive
demands upon the average intelligence, and must be implemented
pluralistically. The guidelines for such a code include that,
1) rules are to be directed at areas of behavior needing
regulation, 2) concrete rules should be made for frequent
situations, particularly ones for which predictability of
behavior is important to many persons (e.g. keeping promises),
3) the rules cannot be too numerous, since the system must be
comprehensible to the average person. The mediation model
suggested here is meant to satisfy these guidelines by
1) regulating speech acts and emotional reactions, 2) imbedding
frequently used rules in automatic mechanisms and 3) justifying
rules on the basis of biosocial needs. The political theory
of group decision making is a limiting condition of the current
model. That is, the current model is not designed to cope with

gaming, coalition formation, and other antagonistic types of
interaction which are typically subject to social welfare
paradoxes (Arrow, 1963). The model can detect when such inter-
actions are taking place, and is designed to guide the group
away from those types of interaction. The model is most suitable
in situations where individuals represent only themselves, and
therefore base their position and statements upon their own
ideas and feelings.

Biosocial base.

In a system for mediated communication it is important to ensure
that key elements of effective communication are not blocked or
distorted in some way by the communication channel. Rice (1981)
summarizes some reports which noted that the nonverbal cues
lost in the use of computer-mail communication, and the increased
rapidity of response, sometimes led to iterative escalations of
misunderstandings. Signals from facial musculature, much of
which is exclusively devoted to the transmission of nonverbal
cues (Rinn, 1984), is, of course, blocked in non-video-confer-
encing systems. Even such systems may distort non-verbal cues,
such as those associated with eye contact, due to technical
limitations of the medium (Bretz, 1983:198). So even our most
sophisticated mediated communication systems have a distorting
effect on communications. That being the case, it is important
to ask what supportive measures can be introduced into such
systems which will ensure that key elements of communication
are effectively transmitted. The objective of such support is
to regulate both emotional tone and cognitive complexity of
messages so that participants can effectively inform and correct
each other's feelings as well as each other's reasoning. We will
concentrate here on the problem of emotional stabilization.

If a mutual emotional stabilization is to be provided both
simply and successfully it is necessary to know the number of
independent dimensions to be controlled. Then channels which
allow the expression of an individual's state on, and control of,

4.

each dimension are necessary.

In a comprehensive overview of the biology of emotion and other
feelings, Pribram (1980) makes a distinction between emotional
and motivational feelings. Emotions are chacterized as internal
control mechanisms which only have an influence beyond the
organism in communicative settings. This results from the ability
of organisms to "read" each other's emotional expressions. The
effectiveness of these responses on the external environment are
therefore entirely dependent upon the sensitivity of the reci-
pient of these expressions.

Emotional states are identified with bodily homeostasis through
chemical regulation and are associated with three centers in the
core of the brain. The frontal intrinsic cortex (amygdala) is
involved in endorphin (morphine-like substance) homeostasis and
appears related to momentary arousal produced by novelty. The
posterior intrinsic cortex (basal ganglia) is involved in dopa-
mine homeostasis and appears related to motivational readiness
and feelings of effectiveness. The hippocampus is involved in
enkephalin and pituitary-adrenal hormone regulation and plays
a role in coordinating the other two systems. "When demands of
arousal are pitted against those of continuing readiness, the
feelings of stress and effort are experienced (Pribram, 1980:257)."
These three systems like the sensors related to hunger and thirst
are sensitive only to amounts, and yield global responses to the
organism's stimulus environment.

The picture which emerges from this analysis is one of a balance
between perturbation (upset) and equilibration (stability) which
is maintained by two pairs of opposing brain systems. The three
homeostatic systems are complemented by the extrinsic cortex
which is responsible for perceptual and motor skills. It is
inputs from the extrinsic cortex which identify feelings with
respect to their qualitative context. Pribram (1980) claims that
three independent dimensions arise from the interaction of these
systems. One, the affective-effective dimension is crucial for
the argument here because it suggests that by analysis of motiv-

ational feelings and related behaviors which influence the en-
vironment, one can infer the affective state of the organism.
These behaviors can be expressed either explicitly in verbal
expression or implicitly through nuances of social interactions.

Implicit communication.

Since the communication systems we are discussing are constructed
to transmit explicit communications, and usually function well in
doing so, a closer examination of implicit communication is needed.
In summarizing his work in this area Mehrabian (1984:18) comments
"Emotions were easily analyzed in terms of three basic and in-
dependent dimensions, pleasure-displeasure, arousal-non-arousal,
and dominance-submissiveness." This referential analysis yields
a pair of bipolar dimensions pleasure-displeasure (earlier called
positivity), and dominance-submissiveness (earlier called potency
or status), and a unipolar dimension, arousal (earlier called
responsiveness). These communicative dimensions seem applicable
both to implicit and explicit communication. Research results
on facial and vocal expression, hand gestures, posture and po-
sition seem to be well accounted for by this approach which
analyzes the referents of implicit communication.

Since the analysis is applicable to both explicit and implicit
communication we can contrast the channels to assess their relative
importance. When there was inconsistancy between the channels, in
face-to-face communication only about 7% of total "feeling" was
carried by explicit verbal expression. Facial expression and
vocal nuance carried the preponderance of emotional tone. When
the channels yielded inconsistent messages the implicit channels
dominated (Mehrabian, 1972). In inconsistent messages the explicit
component conveyed evaluation of addressee's behavior while the
implicit component conveyed evaluation of addressee's person.
In Western culture explicit communication of feelings, particularly
nagative ones, is discouraged and therefore more acceptable in
implicit channels. It is argued that implicit behaviors are more
important in a variety of contexts because they are more difficult

to censor and also may permit greater subtlety of expression.
It was also argued that the three dimensions were useful not only
in characterizing the referents of communication, but also in
describing primary aspects of social behavior. On the basis of
factor analytic studies which did not use a priori categories
of cues in an analysis of social interactions, the data indicated
that positiveness (affiliative behavior), responsiveness (activity
level), and dominance-submissiveness (relaxation) were fundamental.

In summarizing some earlier work Mehrabian (1972) suggests that
communication accuracy increases with the number of channels,
the number of channels available compared to the number habitually
required by the communicator in a given instance, the decoder's
ability to modify the rate of transmission of a message, and
feedback to the communicator about how the message was being
received. In the next section we define some additional channels
for communication and discuss their impact upon communication
accuracy.

Implied validity claims in dialog.

Formal procedures for dialog management, such as parliamentary
procedure, have structures which give priority to statements
which deal with the management of meetings (privileged motions),
and secondarily to questions arising out of procedural consider-
ations (incidental motions). These motions take precedence over
and suspend the discussion of substantive questions. They often
question the validity of certain actions. These formal procedures
and other less formal methods of policy argumentation function by
questioning the assumptions, most often implicit, upon which the
dialog is based.

Habermas (Bernstein, 1978; McCarthy, 1978) has put forward a general theory of
communicative competence which distinguishes between these two
modes of dialog. He defines "communicative action" as speech and
implicit communication which requires a background consensus which
is accepted or taken for granted. "Discourse" is defined as dialog

which occurs when this consensus is disturbed and the validity of
assumptions are called into question. He identifies four irreduc-
ible validity claims which refer to different domains.
1) Comprehensibility is fundamental for understanding and refers
to the domain of language. 2) Truth is fundamental for the develop-
ment of shared knowledge and relates to the domain of "outer nature"
or science.3) Veracity or truthfulness is fundamental to the establish-
ment of trust and relates to the domain of inner nature and inten-
tions.   4) Rightness or legitimacy is fundamental to agreement
and generation of accord, and relates to the domain of social inter-
action. If any one of these validity claims are questioned then
speech action is suspended while discourse takes place in an attempt
to restore the background consensus.

Questioning the validity of claims behind statements in this way
can be seen as explicitly operationalizing the communications which
are normally carried in implicit channels. One justification for
the primacy in importance of implicit communication is its useful-
ness in assessing the truthfulness of speakers. Similarly, feed-
back as to comprehensibility, agreement (rightness) and truth of
statements are most often communicated effectively in face-to-face
interaction without recourse to explicit statements. Agreement can
be seen as following the dimension positivity and matching the
primary social behavior of affiliation. Disbelief can be expressed
either positively, as laughter, or negatively, as disdain. In the
first case, severe misunderstandings can occur when humorous or
ironic statements are taken seriously. In the second, norms against
explicit negative evaluation of a speaker as lying requires support
of this crucial feedback. While these parallels seem compelling,
the particular validity claims listed are not crucial to the
overall argument. What is important is the idea of priority for
such claims and the notion of a discursively achieved consensus.
While Habermas's analysis is based upon contemporary philosophy of
language and theoretical linguistics, Plato develops a similar
one in the Phaedrus (Bernstein, 1978:262). Both argue that if
dialog is to lead to truth, an ideal speech situation is needed
which in turn presupposes an ideal social environment. In the next
section the elements of such an environment for mediate communi-
cation are considered.

A group communication environment.

First, we reviewed an ethical theory which defines rationality in
terms of maximal criticism. It emphasized the importance of rules
in guiding behavior and I suggested that such rules regulate speech
opportunities and emotional reactions. The rules are imbedded in ..
an automatic mechanism and are justified on the basis of biosocial
needs. An analysis of feelings suggested a three dimensional struc-
ture of emotions with a balance between affective and effective
reactions. A review of implicit communication also uncovered three
fundamental dimensions of emotional referents and modes of social
behavior. Finally, an analysis of the implicit validity claims in
speech situations suggested four fundamental challenges that would
be raised in dialog, and suggested a consensus theory of truth
based upon ideal speech and social environments.

While there are significant links and mappings between these areas,
it is beyond the scope of this paper to address them comprehensively.
Some specific links have already been mentioned, so I will here take
a global view of the interrelations. Brandt's ethical theory as-
serted the importance of fully informed and corrected feelings in
approaching rationality. The current inability of mediated com-
munication systems to comprehensively transmit implicit cues which
are felt most important in communicating emotions in an unbiased
manner, suggested that understanding of the dimensions of emotional
response could be crucial. The close coupling of affective and
motivational responses implied that by analysis of motivated (ex-
plicit) behaviors one could assess emotional states. An analysis
of the implicit communications illustrated their importance and
their parallel to the explicit validity claims appropriate in dialog.
Thus, it is implied that by supplying explicit channels for challeng-
ing the validity of statements, we can obtain the information neces-
sary for emotional regulation in the dialog process.

In relation to the dimensionality of emotional response, the analysis
of the biological substrate revealed four poles, three identified
with biochemical homeostasis and the fourth associated with per-
ceptual and motor skills. This system generated three dimensions
with some parallels to the three referential dimensions of implicit

communication, pleasure-displeasure, dominance-submissiveness,
and arousal. Arousal, however, is a unipolar dimension so the
mapping is not complete. The three dimensions can also be used
to characterize explicit communication , so an additional dimension
or pole is required for complete mapping. It is not clear whether
arousal can be complemented with another dimension representing
explicitness of communicative behavior in order to yield a more
perfect mapping (there is some evidence which suggests this).
However, it seems clear that three emotional dimensions are con-
trolled by four behavioral poles. If we assign each of the validity
claims to a separate channel, then we have four expressive modes
which, if the mapping is appropriate, allow the full range of emot-
ional and explicit expressions. I assume here that explicitness
of activity can be captured on a single dimension. This assumption
may not be correct, but an analysis of cognitive dimensions is
beyond the scope of this paper.

A comprehensive model for emotional regulation must have both
measurement and control structures. The expressive modes serve the
function of measurement, since they allow persons to indicate to
the system reactions to others and their statements. Normally,
emotional reactions only affect the environment to the degree that
others are capable of "reading" these reactions and respond to
them. Since mediated communication can distort these signals, a
strategy of explicit response from which emotional states can be
infered has been suggested. This inference process takes place
directly in the mediating mechanism and therefore is not directly
dependent on the ability of persons to "read" emotional responses.
What the mechanism does, is restructure the communication environ-
ment in such a manner that self-management processes regulate
emotional states in the individual and operational point of the
system. I will illustrate how a performance oriented criterion can
be used to find an ideal operating point. For simplicity, let us
assume that control is accomplished by altering the cost of sub-
mitting a criticism to a statement through the four channels which
are available to challenge validity claims. Since this automatically
supplies control dimensions which match the expressive dimensions,
stability is possible in the system.

Having specified dimensions for expression and control, let us
evaluate some evidence from the literature on group decision
making in order to specify the operating point for emotional
stabilization on the three referential dimensions of emotional
response.

Both fact finding and objective appraisal of alternatives is
crucial if groups are to be effective in decision making (Janis,
1972:4). This echoes Brandt's ethical theory analysis which
called for informed and corrrected feelings and ideas as a
criterion for rational decision making. Also of importance is
high group cohesiveness defined as when members express solidarity,
mutual liking, and positive feelings about carrying out group work.
When cohesiveness is high, there is greater participation, less
anxiety, and heightened self-esteem. However, pressure toward
conformity also increases under these conditions. Deviant members
at first become the target of increased communication and if they
do not change their dissident ideas, are excluded, first subtly
and then in more obvious ways, in order to restore unity of the
group. An informal objective to preserve friendly intragroup
relations can result in concurrence-seeking and striving for
unanimity which can override motivation to realistically appraise
alternative courses of action. Janis (1972:9) refers to this as
"groupthink", "a deterioration of mental efficiency, reality test-
ing, and moral judgement that results from in-group pressures."
The opposite extreme which Janis (1972) identifies as increasing
the probability of error in decision making, is deliberate con-
formity out of fear of recrimination. In both cases there is a
lack of criticism and disengagement from the realities of the
situation, with "groupthink" it results from overoptimism and
lack of vigilance, and in the other case which might be called
apathetic withdrawal, it results from fear. In terms of the
referential dimensions of emotional response, one extreme is ex-
cessively pleasurable and the other excessively displeasurable.
However, both could be characterized as being excessively sub-
missive or lacking potency as far as most members are concerned.

11.

The opposite extreme on the dominance-submissiveness dimension
which can be characterized as chaotic argument, can also lead to
errors in decision making: Bitter power struggles can occur when
members have no sense of loyalty to the group and regard them-
selves merely as representatives of groups with conflicting
interests. Excessive dominance by a few group members may force ..
others into excessive submissiveness and reduced participation,
and perhaps apathetic withdrawal. Either situation is likely to
yield displeasure for most members when dominance is directed
towards those in the group. When directed against out-groups
excessive dominance may yield a type of groupthink characterized
by pleasurable feelings and authorization of dehumanizing actions
against out-groups. Janis (1972:12) calls this "hardhearted actions
by softheaded groups".

This type of analysis supplements the rational actor theory of
individual decision making with a group dynamics approach to
explain how errors of individuals can be augmented by group
processes.

It appears, therefore, that either extreme of the pleasure-
displeasure dimension can lead to errors in decision making.
In the above mentioned case, group feelings are so positive that
the negative feelings generated by the contemplation of dehuman-
izing actions are overwhelmed by the predominant positive tone.
Feelings of elation can be seen as blocking the process of fully
informing and correcting feelings which is required by Brandt
(1979) for reaching rational and ethical decisions. Extremes
of dominance and submissiveness can result in poor participation
and a lack of information being made available to the group by
its members. It can be seen as disrupting the process of criti-
cism by facts and logic necessary for avoiding reasoning errors
(Brandt, 1979). One extreme is too much internal criticism where
no one admits to errors and the other is too little criticism
where erroneous ideas remain unchallenged. Such disruption of
the critical process also makes it hard for individuals to assess
each other's competence on the matter in question. It is the
regulation of this process which will be examined in detail later.

On the basis of the above, we can make some suggestions about the
ideal operating point in individual's emotional space. On the
pleasure-displeasure dimension considerations of group cohesive-
ness require pleasurable feelings in group interaction. However,
these feelings should not be far from the neutral point in order to
permit negative emotional reactions to inform the group's decisions.
On the dominance-submissiveness dimension the ideal point is slight-
ly toward dominance. This would be expected to generate criticism
when warranted and also some competition for speaking turns. The
negative effects of excessive dominance have been mentioned above.
The ideal point in terms of speaking turns would be achieved when
one, and only one, person wished to speak at each transition.
Under this unrealistic condition no conflicting requests to speak
would need resolution. A more realistic situation is to generate
enough requests so there is always a speaker available and con-
flicting requests are automatically resolved so the conflict does
not consume group resources. Arousal levels would best be regul-
ated with reference to two objectives. Since higher levels of
arousal strengthen emotional feelings, arousal should be reduced
if regulation becomes unstable and deviates from the ideal point.
The second objective relates arousal levels to cognitive com-
plexity. Higher arousal results in a narrowing of attention,
reduced encoding of cues, and increased forgetting (Christiansen,
1984). Thus, as more parallel or shared processing is required,
arousal levels should be reduced. Too much reduction could, how-
ever, result either in boredom or sleep on one hand, or excess
attentional resources which are available for gaming. It is most
probable that changes on any one of these dimensions would result
in changes on the others. We could also expect changes and inter-
actions over time as a group moves through different stages of
self organization and problem solving. These complications are
beyond the scope of this paper. The focus here will be on uniform
stability in only the level of dominance as manifested in criticism.

Finally, let us set the stage for a more in depth analysis of
the regulation of criticism by describing the options available
to a person in the hypothesised teleconferencing situation.
Earlier, I listed the four validity claims thought to be implicit

in any speech act. If questioned, the challenge to that claim
took priority in the dialog process. Such discourse continued
until the challenge was resolved, thus permitting speech acts to
occur once again. Thus a clear priority between discourse and
speech acts is specified. Within discourse some priorities appear
obvious. Comprehension must occur before assesssment can be made ..
of a statement. Similarly, if a statement is meant in jest or
felt to be untruthful, it makes little sense to evaluate it in
any other way. Next level of priority could be assigned to claims
that a statement is true in its propositional content, either
factually or as a logical conclusion from earlier accepted speech
acts. Finally, the question of whether a statement is right in
its performative content, do others agree that the specified acts
are legitimate. If we accept this ranking then we can prioritize
criticisms so that challenges to a given statement are dealt with
in the order - comprehensible, truthful, true, and right. Since
these tend toward mutual exclusion, the buttons in front of each
speaker might better read "don't understand", "not serious",
"false", and "disagree". If a person did not press one of these,
the acceptance of the validity claims would be assumed.

Within parliamentary procedure different motions, depending on
their priority level, have different effects on the dialog process.
Those of highest priority may interrupt the current speaker. Under
normal circumstances such an interruption could be generated by a
person indicating lack of understanding of the current speaker using
the system proposed here. That person would be expected to explain
the reason for the interruption immediately. The "not serious"
response could also be treated in a special manner. It could also
interrupt a statement. Other offers of criticism, triggered by
"false" or "disagree" indication could be processed at the com-
pletion of previous statements. The effect of this prioritizing
would be to select those on opposing sides of an issue to speak on
alternating turns. This type of balance has been found to lead to
more objective evaluation of issues. The ability of certain responses
to cause interruptions could be selected according to the purpose
of group. In a learning environment interruption might be very
limited.

14.

In a performance oriented group operating under time pressures,
interruption might be permitted in response to any challenge of
a validity claim.

In the category of speech acts we can also imagine a hierarchy
of choices. Highest priority might go to an immediately related
comment which builds upon the previous statement. Next, we might
want short comments, followed by just plain statements. Finally,
statements relating to a new topic might receive lowest priority.
These buttons could read - "constructive comment", "short comment",
"statement", and "new topic". These suggestions go beyond the
question of selecting the next speaker to the question of manag-
ing the flow of topics in discussion, and are therefore beyond
the scope of this paper. Let us select the challenge to the truth
claim of a statement from the category of discourse, and for sim-
plicity limit criticism to the logic of a statement. From the
speech acts category we will select "statement". Criticisms then
will always have priority over statements.

Within a given priority level, conflicting requests are resolved
in favor of the person who has thus far spoken the least. We
assume that people wish to maximize their speaking time. This
assumption is based upon two well known psychological theories.
The first is derived from the well supported finding that degree
of opinion change is influenced by the number of messages received.
Thus a speaker is more likely to persuade others to support the
speaker's position if the speaker has more time to present messages
to them. The second is the social comparison theory which is based
upon the assumption that people wish to find out if their ideas
are correct, and will do so by comparing them with those of similar
others (Deutsch & Kraus, 1965). Validation of one's ideas is
assumed to in turn promote increased participation.

A theory of criticism.

The purpose of this section is to outline a rigorous theory of
responsible criticism applicable to automatically mediated dialog.

Critical thinking is an active cognitive process in which listeners
attend carefully to presented ideas and try to match those ideas
with their own memory structures. Responsible criticism occurs if
a reasonable attempt to match presented ideas against those pre-
viously held fails and produces a comment which is directed toward
correcting a flaw in the generating memory structure.

The underlying assumption in the construction of the theory is
that of impartiality or equal opportunity for all persons. Three
basic elements of the model are the psychological categories of
stimulus, response, and reinforcement. Two types of vocalizations
serve as stimuli; they are denoted "Statement" and "Criticism".
Each of these is defined according to whether it follows the
response of pressing "Statement" or "False" button. The button
response is considered to be a covert or incipient vocalization.
A response is reinforced when the associated vocalization generates
no critical comments. The purpose of reinforcement is to associate
vocalizations and responses. The processes of speaker selection
and error verification serve to generate responsible criticism.
 Participants in the dialog are assumed to be rational persons
simultaneously focusing their attention upon current statement.
These assumptions are necessary in order to achieve mutual knowledge
essential for effective communication (Clark and Marshall, 1978).

In social interaction, accountability, or more precisely its
expectation, precedes responsible action. In automatically mediated
dialog, one mechanism of accountability is a speaker selection pro-
cedure. Persons indicating that an error in reasoning has been made
increase their probability of being selected to speak and being
required to identify the flaw in the previous statement. This
possibility could dampen tendencies toward being overcritical. If
the person could not identify the error, the person's estimated
probability of correct "Criticism" would be reduced. This would
reduce their influence on group process. Thus vocal justification of
error indication is one mechanism of accountability.

This accountability structure necessitates the identification and
verification of errors in statements and in criticisms. Case 0

(Figure 1) indicates notation which precedes and follows error
processing. The simplest identification-verification sequence
occurs when a statement by person A triggers a "False" indication
by person B (Case 1). B is then selected to speak, accurately
criticizes A, and the criticism is verified (accepted) by all
participants, since none press their "False" buttons. In this
case, A is charged with one statement (S^-) error and B is credited
a correct criticism (C^+). In general, agreement (S^+) precedes
and terminates an error processing sequence.

A double error sequence (Figure 1, Case 2) occurs when B's criticism
is found to be in error by C who accurately criticizes B. A simple
error processing algorithm would charge B with an erroneous criti-
cism (C^-) and credit C with a correct criticism (C^+). A more
sophisticated scheme would permit C to then criticize A.

A possible problem in processing errors is the sequence in Figure 1,
Case 4. Here A and B speak repeatedly and each indicates the other's
criticisms are in error. This type of recursive malfunction might
be termed an "argument". The rule illustrated (Figure 1, Case 4)
would terminate the sequence, after giving each speaker two criti-
cisms, by selecting a new speaker, in this case C. A would be
charged with an unverified statement (S^0) and two unverified criti-
cisms (C^0).

A simple approach to evaluating each verbalization is to compute
its contribution to the probability of being correct, where correct
is defined as a verbalization generating no "False" indications.
Thus S^+ is assigned the value of one and S^- is assigned the value
zero, for each of N statements made by a given person. If we con-
sider only verified statements, a person's estimated probability
of making a correct statement is:

$$P(S^+) = \Sigma S^+/N$$

A similar score is derived for verified criticisms.

This analysis has been simplified by considering only the contri-
bution of verified vocalizations to the person's estimated

probability of being correct. Button responses can also be used
to estimate this probability. First, consider the four possible
outcomes during each statement (Table 1). The "No Error" condition
is represented by Case 0 (Figure 1); all participants are credited
uniformly where everyone agrees statements are correct. The "Hit"
is illustrated in Case 1 where B correctly criticizes A. Since
other persons did not offer critical comment, they are charged
with a "Miss". A person's verbalization and button response can
be scored independently with this approach. "False Alarms" cannot
be derived from knowledge of statement scoring with certainty;
they are a subset of the responses to the unverified statements
(S^0). In case 2, B's criticism was erroneous, either because A was
correct ("False Alarm" by B), or because it did not make sense as a
statement ("Hit" by B). Thus, only an upper limit can be given for
"False Alarms" in terms of the scores established for verbalizations.
Likewise, only the lower limit for "Hits" can be so derived.

When consensus of no error is used to verify statements, a certain
proportion remains ambiguous, because a consensus is not always
reached quickly enough to verify statements. These unverified
statements can be estimated as either correct or in error by
evaluating the button pattern (BP) of the group and the prior
probability of a correct statement $P(S^+)$ by that speaker. If we
know the probability of a button pattern occuring given that a
statement is true $P(BP|S^+)$, then by Bayes' Theorem we can determine
the probability of a true statement given the button pattern
$P(S^+|BP)$.

$$P(S^+|BP) = \frac{P(BP|S^+)\ P(S^+)}{P(BP|S^+)\ P(S^+)\ +\ P(BP|S^-)\ P(S^-)}$$

If the button pattern (BP) is, for example, not "False", not
"False", "False", "False" (nnff), then

$$P(S^+|nnff) = \frac{P_1 P_2 (1-P_3)\ (1-P_4)\ P(S^+)}{P_1 P_2 (1-P_3)(1-P_4)P(S^+)\ +\ (1-Q_1)(1-Q_2)\ Q_3 Q_4 P(S^-)}$$

18.

where

$P_i = P(n|S^+)$ = probability of not "False" to a correct statement by person i ("No error")

$Q_i = P(f|S^-)$ = probability of "False" to an erroneous statement by person i ("Hit")

This rule is used to adjust the prior probability of a statement from that person being correct. Similarly, the probability of a keyboard response being correct, can be updated by Bayes' Theorem. Thus, inability to validate each verbalization need not prevent the scoring of each statement or criticism.

When the amount of criticism increases, a large number of unverified statements and criticisms could be expected. While it is not possible to verify them if processing is limited to the linear process described above, it is possible if some type of storage mechanism is introduced. If each statement is digitally recorded, then a "False" indication can automatically result in its transfer to a push-down stack. Thus, if criticisms immediately follow each other these challenges can be processed with the most recent one resolved first. After agreement has been reached on the most recent criticism, the previous one could be "popped" off the stack and played back by the computer. If new criticisms were triggered in this process, additional statements would be placed on the top of the stack. How completely this process would be applied, could be determined by the type of group or it could be triggered by a level of stress jeopardizing the emotional stability of the group.

The concept of stress can be operationalized and quantified within the present framework. In the situation where it is either not timely or not possible to verify the correctness or incorrectness of each statement or criticism, there may be a substantial deviation between the model parameters for a person which have been verified by the group consensus and those which have been estimated by the model. Large differences would indicate an individual deviating significantly from the group consensus and subject to increased stress.

Table 1

Statement-button press matrix for dialog support
challenging the validity claim "true"

Button Press

	"False"	not "False"
False	Hit 1	Miss -1
True	False alarm -2	No error 0

Statement

P (Hit) + P (Miss) = 1

P (False alarm) + P (No error) = 1

Key

Name of
event

        Pay off
      in minutes
   removed from
accumulated speaking
time

Figure 1

Possible outcomes of linear error processing

Case 0

	A	B	C
Vocalization			
"False" indication			
Scoring	s^+	s^+	s^+

Case 1

	A	B	C
Vocalization			
"False" indication			
Scoring	s^-	c^+	s^+

Case 2

	A	B	C	D
Vocalization				
"False indication				
Scoring	s^{0bc}	c^{-c}	c^+	s^+

Case 3

	A	B	C	D	A
Vocalization					
"False" indication					
Scoring	s^{0b}	c^{0c}	c^{-d}	c^+	s^+

Case 4

	A	B	A	B	A	C
Vocalization						
"False" indication						
Scoring	s^{0b}	c^{0a}	c^{0b}	c^{0a}	c^{0b}	$s+$

Time ——————→

Participants = A,B,C,D

Key: s^+ verified correct statement

s^0 unverified statement

s^- verified incorrect statement

c^+ verified correct criticism

c^0 unverified criticism

c^- verified incorrect criticism

21.

Dynamic regulation.

In purely rational terms, the ideal operating point for a group
is one in which criticism is maximized. By maximal criticism we
reduce the number of times the group comes to an explicit con-
sensus. The model developed in the previous section can be used
to "guess" how a given challenge would have been resolved. This
creates some "stress", but permits the group to maintain a higher
level of criticism and to move forward more rapidly.

In table 1, each event has associated with it a payoff given in
minutes. Thus, if a person indicates "False" to an erroneous state-
ment and is supported by the group consensus, a "Hit" is said to
occur. With the nominal payoffs shown, the person would get one
minute removed from accumulated speaking time for each "Hit".
If "False" were not indicated and a statement was shown to be
erroneous, then one minute would be charged to the person's
account. This is called a "Miss". Thus a person not attending
would easily lose the opportunity to speak in the future in com-
petition with others. If "False" was indicated in response to a
statement which was agreed to be true, the person so indicating
would be charged two minutes for that "False Alarm". Finally,
if a correct statement were made and no "False" were indicated,
the "No Error" condition, no time would be removed from or charged
to a person's account. Given that a statement is correct, a person
must be charged for either a "False Alarm" or a "No Error".
Similarly, if a statement is false, then a person will be charged
for either a "Hit" or a "Miss".

By adjusting the payoff for a "Hit", the level of criticism can
be regulated. This payoff must be increased to generate more
criticism and decreased to reduce the level of criticism. This
balance can be just as well affected by adjusting any of the
payoffs, so the control system has a single degree of freedom.
The strategy for adjusting the payoff for "Hits" is presented
in Figure 2. When a statement is challenged, the computer cal-
culates the probability that the group will support the criticism.

This calculation uses Bayes' Theorem and combines historical
information which gives the probability that the individual
would make an erroneous statement (available before the state-
ment is heard), with the indications of "False" or not "False"
from listeners, which when their "Hit" and "No Error" probabilities
are known, can be used to estimate the outcome of the challenge
(before the criticism is heard). If the probability of a success-
ful challenge is greater than .5, the machine guesses that it
will be sustained. If it is right, then the payoff for "Hits" is
increased. If the guess is wrong, the payoff for "Hits" is de-
creased. The amount of increase or decrease controls how stable
the estimates are and therefore how rapidly the model can adjust
to changes in topic or other circumstances which alter relative
competences.

At least two situations could result in incorrect guessing by the
model. The first is a change in the topic of discussion which
alters the relative balance of competence in the group. Under
this condition more verified statements are needed so the model
can be recalibrated. Lowering the payoff for "Hits" has the effect
of limiting the number of unverified statements. The second situ-
ation relates to a basic assumption of the model. The probabilities
are assumed to be constant, or very slowly changing. If, for instance,
a person attempts to game the situation by saying in effect "I know
this statement is wrong, but I will say it anyway, because no one
will know and I can escape criticism," this assumption of the
model is violated. This situation in which probability of a correct
statement is not stable would cause the model to guess inaccurately
and again reduce the payoff for "Hits". This reduction would move
the group toward greater concurrence-seeking, which tends to
increase group cohesiveness at the risk of "groupthink". The in-
creased solidarity, mutual liking, and positive feelings about
carrying out group work would be a powerful antidote to gaming
and other antagonistic tendencies. Therefore, while the model can
not cope with gaming, it can detect when it may be occuring and
guide the group to a state in which feelings of solidarity dominate
the decision making activity. Thus, social welfare paradoxes can
be avoided by taking advantage of human needs for affiliation and
self-esteem.

23.

An alternative payoff system could be based upon a "No Error"
condition resulting in a charge approximating the mean time
for a statement. This would give participants a more realistic
assessment of the time cost associated with listening to a
statement. Another possibility would be to relate the payoffs to
the actual time used by a given statement. Thus a "Hit" resulting
from a criticism to a long statement would be worth more than one
resulting from a criticism to a short one. This could have the
result that a very short statement might not be criticized,
because the time used to state the criticism would actually cost
more than the payoff from a "Hit". Within a given group only the
relative balance between individual time accumulations has an
effect upon speaker selection. Thus only the relative costs of
statements and criticisms are meaningful. If persons could trans-
fer their time accumulations to conferences with new groups, ab-
solute levels of time charges could be meaningful. The various
alternatives proposed here might be useful in different kinds
of groups or might apply to different challenges in the same
group. For instance, a short humurous comment would most often
not need a challenge, even though it violated the background
consensus by being "Not Serious".

 Analysis of prejudice and stability.

The analysis thus far assumed that individuals can adjust
optimally to the payoffs and that they are equally critical of
statements regardless of the speaker. However, some persons
are prone to be overcritical and others often decline to offer
criticism even when warranted. Another problem results from the
tendency to be more critical of some group members than others.
Two analytic techniques can be applied to separate tendencies
toward criticisms resulting from actually perceiving errors and
those resulting from prejudice. Whenever a statement is made,
it can generate a criticism indication. From these indications
two sets of data can be derived. The first is a matrix which
indicates the probability that a given person will offer criti-
cism when a given other makes a statement. These indications can

be strongly influenced by communicator attractiveness or other
factors prejudicing the listener to lower or higher levels of
criticism toward a given other. A second half-matrix of probab-
ility data can also be derived which shows pairwise co-variation
among individuals. These data are not directly influenced by
prejudice which may exist between the two listeners, since they
are not aware of each others' responses. However, the degree of
co-variation may be influenced by the speaker. That is, some
speakers may generate more dissensus in the responses of their
listeners than others. It may therefore be useful to think of
each speaker as associated with a vector of responses from the
previously mentioned full matrix and a half-matrix of co-variation
data. The mean level of agreement between each pair of persons in
the co-variation matrices can be compared to each person's agree-
ment with the other's statements. The contrast between reaction
to public and anonymous responses of the other person can yield
a measure of prejudice.

The co-variation data can also be used to estimate the stability
of the group. The theory of cross-cutting coalitions states that
group stability is high if different issues split the group into
relatively equally balanced coalitions which are different for
each issue. Thus, depending upon the issue, a person will have
different sub.-groups of persons who support and oppose him. If
we assume that each person's position represents an "issue",
then the relative balance of agreement and disagreement among
the listeners, and the cross-cutting of these sub-groups which
are generated by each speaker, can be used to estimate group
stability. This estimate could be used to indicate the degree
of stress a given group could tolerate before disintegrating.

A second method for bias estimation uses signal detection theory
extended to deal with identification of multiple signals (Swets
et al., 1978). The theory permits an individual's performance
to be analyzed in terms of two independent parameters. The first,
sensitivity, is a measure of the person's ability to detect and
identify an erroneous statement. Sensitivity is defined as the
probability of a "Hit", less the probability of a "False Alarm".

$$\text{Sensitivity} = P(\text{Hit}) - P(\text{False Alarm})$$
$$\text{Bias} = P(\text{Hit}) + P(\text{False Alarm})$$

The second, bias, is the sum of the same two probabilities. Bias
can be thought of as the tendency to criticize assessed inde-
pendently of ability to detect and identify erroneous statements.
Sensitivity is assumed to be fixed by ability, but bias can change
both with changes in payoffs and depending upon who is speaking.
A measure of prejudice can be derived by comparing the person's
ideal bias point, given the payoffs, and the actual bias as
measured. The computation of these parameters can be imprecise
as in Figure 1, case 2. Only an upper limit can be given for
"False Alarms". Likewise, only a lower limit for "Hits" can be
so derived.

The inability to derive these probabilities with certainty is
not a serious problem for the signal detection model because
one parameter is relatively insensitive to this inability. The
bias parameter, indicating the probability to indicate error,
is incremented by either a "Hit" or a "False Alarm".

Thus, when a statement is unverified, we can say that those in-
dicating error are more critical than those who do not. This lack
of verification, obviously, does not permit us to say anything
about a person's sensitivity to error. One assumption of our model
is that sensitivity is relatively fixed; thus the inability to
derive "False Alarms" and "Hit" probabilities precisely from scoring
of statements is not a serious one.

Every challenge which is verified generates sensitivity data from
each person, which can by using this model be integrated. Bias
parameters, on the other hand, require separate estimation de-
pending on the current speaker in order to account for prejudice.
Thus, the lack of dependency upon verification is extremely help-
ful in bias parameter estimation.

Multi-level regulation for overall stability.

In the first part of this paper we discussed implied validity
claims and then showed how they could be used to develop additional
(anonymous) channels for communication in a teleconferencing en-
vironment. We then elaborated a theory of criticism, a model for ..
dynamic regulation, and a detailed data analysis scheme applicable
to a single anonymous response channel. The earlier described
communication environment, however, specified four of those channels
and at least one for "statements".

It is beyond the scope of this paper to attempt a comprehensive
integration of expressive channels with a control structure de-
signed to stabilize groups and individuals as they move through
the problem solving process. However, we can attempt to character-
ize groups according to their payoffs for "Hits" which control
the level of challenges to each validity claim. If, for instance,
a group has stabilized with high payoffs for "Hits" on all channels,
then we can assume that a background consensus has developed which
permits speech acts to take place with little interruption. This
would be indicative of a serious, well organized group.

If a group has high payoffs on all channels except "Disagree", we
can assume there is a consensus on facts and logical procedures.
This group would be having a problem establishing a consensus
on what is the right course of action from a social standpoint.
They could be informing and correcting each others' feelings.
Similarly, a group showing low "Hit" payoffs on the "False" channel
would not yet have reached consensus on the facts and logic of the
problem situation. They could be informing and correcting each
others' facts and logic. An extreme case of this is where cor-
recting has been turned off (zero payoffs for hits in "Disagreement"
and "False" channels) in favor of information collection. This is
often called "brain storming". All these activities can be charact-
erized as oriented toward learning about and optimizing the group's
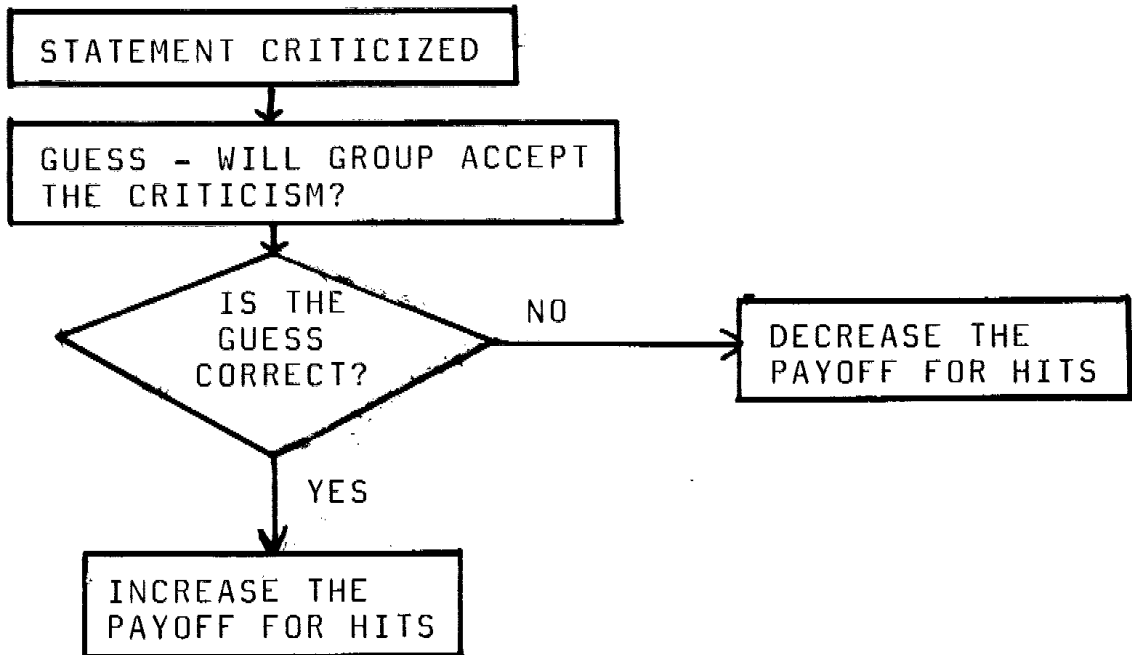response to the environment.

However, in the case of group with a low "Hit" on the "Not Serious" channel, there is a more serious lack of consensus. Either they fail to realize as a group when someone is joking (can't agree about what is funny), or worse yet, doubt each other's veracity. In the most extreme case, the "Hit" payoff for the "Don't Understand" channel drops, making it impossible or worthless for the interaction to continue. This is a clear sign that disintegration of the group may be imminent. In these last two cases the group has become preoccupied with internal conflicts which must be resolved before they can orient themselves to the environment. Groups needing emotional stabilization would be expected to operate in this mode and might even be encouraged to by appropriate payoffs.

Over the course of a group's history, shifts from one mode of operation to another could be expected. A group of diverse individuals might have trouble understanding each other at first, due to use of different languages (this could include problems with professional jargon). Then they might try to agree on how to proceed. Consensus on procedure would then lead to discussion as to legitimate goals for the group. If stress or tension began to impede their progress, a few jokes might be encouraged. The measures of stress and stability combined with inferred emotional states derived from both public and anonymous responses could be used to control payoffs directing the group through these different modes of action.

While the details of such a control strategy are not yet developed, the proposed communication environment provides adequate facilities for the measurement and regulation of behavior in group problem solving. Better models of effective group problem solving methods and individuals' biosocial needs will dictate more comprehensive control strategies for dialog self-management.

28.

Figure 2.

STABILIZING THE LEVEL OF CRITICISM

```
┌─────────────────────────────┐
│  STATEMENT CRITICIZED       │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│  GUESS - WILL GROUP ACCEPT  │
│  THE CRITICISM?             │
└─────────────────────────────┘
              │
              ▼
           ╱IS THE╲          NO      ┌──────────────────┐
          ╱ GUESS  ╲─────────────────▶│  DECREASE THE    │
          ╲CORRECT?╱                  │  PAYOFF FOR HITS │
           ╲_____╱                   └──────────────────┘
              │
              │ YES
              ▼
      ┌──────────────────┐
      │  INCREASE THE    │
      │  PAYOFF FOR HITS │
      └──────────────────┘
```

Self-management of criticism in dialog

REFERENCES

Arrow, K.J. Social Choice and Individual Values. 2nd ed.
        New Haven: Yale University Press, 1963.

Bernstein, R.J. The Reconstruction of Social and Political Theory.
        Philadelphia: University of Pennsylvania Press, 1978.

Brandt, R. A Theory of the Good and the Right. Oxford: Clarendon,
        1979.

Bretz, R. with Schmidbauer, M. Media for Interactive Communication.
        Beverly Hills: Sage Publications, 1983.

Clark, H.L. & Marshall, C. "Reference Diaries." In D.L. Waltz (ed.),
        Theoretical Issues in Natural Language Processing-2.
        New York: Association for Computing Machinery, 1978.

Christianson, S. "The relation between induced emotional arousal
        and amnesia." Scandinavian Journal of Psychology, 1984,
        25, 147-60,

Deutsch, M. and Kraus, R.M. Theories in Social Psychology. New York:
        Basic Books, 1965.

Janis, I.L. Victims of Groupthink. Boston: Houghton Mifflin, 1972.

McCarthy, T. The Critical Theory of Jürgen Habermas. London:
        Hutchinson, 1978.

Mehrabian, A. Nonverbal Communication. Chicago, Il.: Aldine-
        Atherton, 1972.

Mehrabian, A. "This week's citation classics." Current Contents-
        Social and Behavioral Sciences, 1984, 16(43):18.

Pribram, K.H. "The biology of emotions and other feelings."
        Pp. 245-70 in R. Plutchik and H. Kellerman (eds.),
        Emotion: Theory, Research, and Experience. New York:
        Academic Press, 1980.

Rice, R.E. "The impacts of computer-mediated organizational and
        interpersonal communication." Pp. 221-249 in M. Williams
        (ed.), Annual Review of Information Science and
        Technology 15. White Plains, New York: Knowledge
        Industry Publications, 1981.

30.

Rinn, W.E. "The neuropsychology of facial expression: A review
       of the neurological and psychological mechanisms for
       producing facial expressions." Psychological Bulletin,
       1984, 95, 52-77.

Stodolsky, D. Tutorial Dialogs with Automatic Mediation, 1980.
       Unpublished.

Stodolsky, D. "Automatic mediation in group problem solving."
       Behavior Research Methods & Instrumentation, 1981,
       13, 235-42.

Swets, J.A., Green, D.M., Getty, D.J. & Swets, J.B. "Signal
       detection and identification at successive stages
       of observation." Perception and Psychophysics,
       1978, 23:275-89.