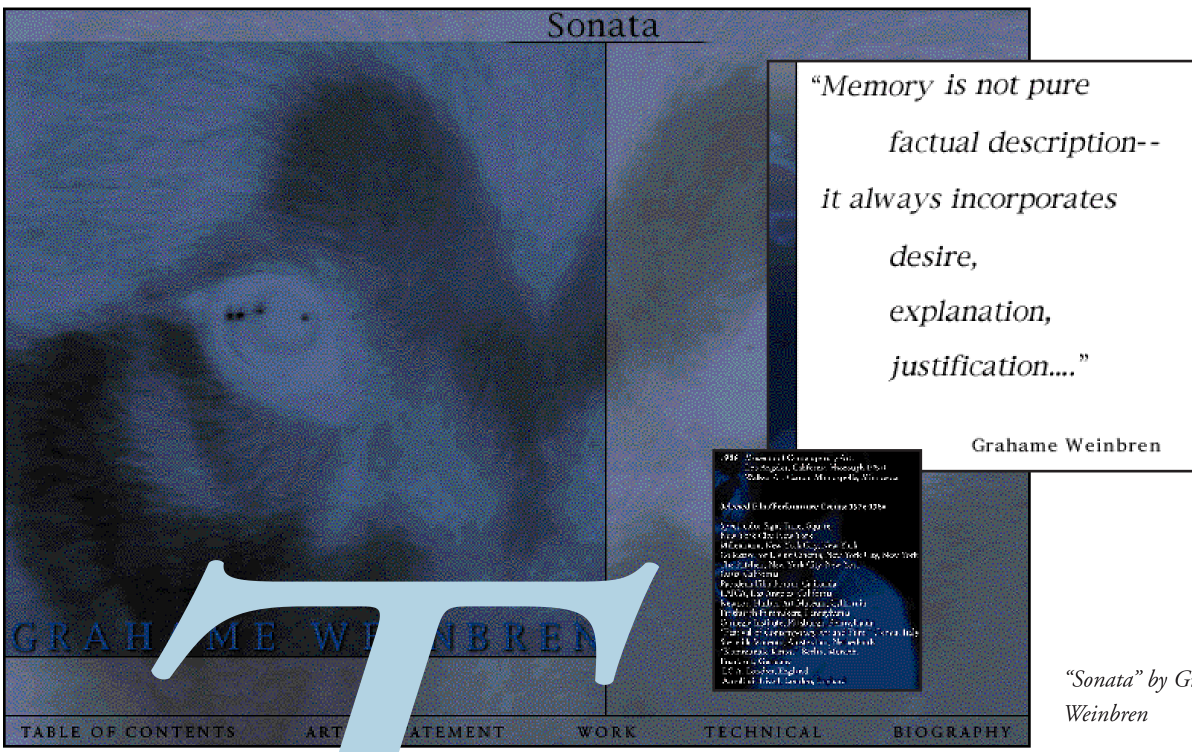
GUY A. BOY

# The Group Elicitation Method for Participatory Design and Usability Testing




*"Sonata" by Grahame Weinbren*

This article presents the Group Elicitation Method (GEM), a brainstorming technique augmented by a decision support system for participatory design and usability testing. GEM has been successfully used in four industrial projects to elicit knowledge from users, management, and designers. In three cases GEM was used to elicit end-users' knowledge for the design of new user interfaces and complex human–machine systems. The properties of this method and the lessons learned are discussed.

**Guy A. Boy**
**European Institute of Cognitive Sciences and Engineering (EURISCO)**
**4, Avenue Edouard Belin, 31400 Toulouse, France**
**+33-62-17-38-38**
**boy@onecert.fr**

METHODS & TOOLS
COLUMN EDITOR
**Michael Muller**
*US West Technologies*
*4001 Discovery Drive*
*Boulder, CO 80303*
*+1-303-541-6564*
*michael@advtech.uswest.com*

## Introduction

In industrial environments, several experts may work together to investigate appropriate solutions for a design or usability issue. A critical problem is to derive an acceptable consensus[1] from a group of experts who share neither the same background nor the same objectives. It is not uncommon to find that experts do not understand each other. Strong personalities may dominate a meeting even if they do not contribute much to content. In addition, power differences exist between management and labor. It often follows that group decision making is not always democratic.

The Group Elicitation Method (GEM) was designed to moderate these effects. Six to 10 domain experts (users, managers, and designers) are selected to participate in a session using the GEM process. Experience suggests that the optimal number of participants for a productive session in a reasonable time frame is seven.

A typical GEM session takes a full day and consists of six phases conducted by a knowledge elicitation facilitator:

- Phase 1: Formulation of issue statements and selection of participants. We found that the type of checklist proposed by Nielsen et al. [9] for semistructured interviews was a good starting point for formulating issue statements. GEM currently uses the following checklist as a starting point:
  - ✓ What is the goal of the engineered system that we plan to design or evaluate?
  - ✓ How is the system or its equivalent being used (current practice, observed human errors)?
  - ✓ How would you use this system (users' requirements)?
  - ✓ What do you expect will happen if the corresponding design is implemented (e.g., productivity, aesthetics, quality of work product, quality of work life, and safety issues)?
  - ✓ How about doing the work this way (naive or provocative suggestions)?
  - ✓ What constraints do you foresee (pragmatic investigation of the work environment)?

- Phase 2: Generation of viewpoints. A viewpoint is a situated statement provided by a participant. A premise of this phase, a written brainstorming technique also known as brainwriting[2], was introduced by Warfield [13] to facilitate the generation of ideas and was further developed by Boy [3, 4]. Brainwriting process usually involves passing around sheets that are incrementally filled in by the participants. Each participant writes either a new viewpoint or reacts to viewpoints already written by another participant. For seven participants this phase usually takes 1 hour.

- Phase 3: Reformulation of viewpoints into more elaborate concepts. This phase may take between 1 and 2 hours. It involves four types of operations that correspond to the concept-clustering mechanisms that were described and used in Fisher's COBWEB system [5]:
  - ✓ Classifying the viewpoint in terms of an existing concept (a class of viewpoints),
  - ✓ Creating a new concept,
  - ✓ Combining two concepts into a single concept, and
  - ✓ Dividing a concept into several concepts.

- Phase 4: Generation of relationships between these concepts. Each participant

1 A consensus connotes general agreement. In this article, by consensus I mean the measure of the extent of consensus that is represented by the sum of the triangular matrices generated by each participant. See footnote 3 for a definition of "triangular matrix."

2 The brainwriting technique was introduced more than two decades ago to facilitate the generation of ideas or viewpoints by a group of people [13]. Brainwriting can be used to stimulate a group of experts with the goal of silently expressing their expertise on a precise issue [3]. It enables a group of experts to construct a written shared memory. A person takes a sheet of paper and reads the issue to be investigated. She then adds several viewpoints and puts the paper back on the table, where the set of papers constitutes a shared memory of the meeting. The process of choosing a piece of paper, reading, writing viewpoints, and replacing the paper on the table is continued until each person has seen and filled in all the papers. Thus, each person is continually confronted with the viewpoints of the others and can react by offering a critique or new viewpoints. Generally, a considerable number of viewpoints can be collected with this procedure.

is asked to provide opinions on the relative priorities of these concepts. He fills out a triangular matrix[3] presenting the concepts in rows and columns. Basically, each matrix "box" is filled by a score that can be +1 / 0 / -1 if the row concept is more / equally / less important than the column concept. This phase takes about 30 minutes.

- Phase 5: Derivation of a consensus. A consensus is derived using an interactive computer program. The global score matrix is the sum of all the matrices generated by all participants. To each global score is attached a standard deviation measuring the interparticipant consistency of the global score of the relationship. This phase leads to the expression of a consensus among the participants, which is expressed using four typical parameters: mean priority, interparticipant consistency, mean priority deviation (stability of a concept), and global consensus.[4]

- Phase 6: Critical analysis of the results. We observed that a critical analysis of the generated concept network is constructive and reinforces the consensus. Experience shows that this phase may take between 30 minutes and 1 hour. At the end of the session, a report is prepared by the knowledge elicitation facilitator.

**GEM Properties and Lessons Learned**

Discussion of the validity of GEM is based on 24 sessions involving an average of seven participants per test. The sessions were conducted with experts from a wide range of aeronautics-related domains. During the last six sessions, the participants completed a questionnaire that included the following questions.

3 A triangular matrix expresses the fact that relationships of priority between concepts are symmetrical, that is, participants need to fill in only the upper triangular part of the priority matrix. In practice, each participant is asked to provide his opinion on the relative priorities of these concepts. He fills out a triangular matrix presenting the concepts in rows and columns. Basically, each matrix "box" contains a score. For example, the participant starts at Row 1, Box 2; if Concept 1 is more / equally / less important than Concept 2, he writes +1 / 0 / -1 in the corresponding box, and so on (see Table 1).

Table 1. Example of a Priority Matrix Generated by a Participant

| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -1 | -1 | -1 | -1 | 0 | -1 | -1 | -1 | 0 | 0 | -1 | 0 | 0 |
| 2 | | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 0 | 0 | -1 | -1 | 0 |
| 3 | | | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| 4 | | | | -1 | 1 | -1 | 0 | -1 | 1 | 0 | -1 | -1 | 0 |
| 5 | | | | | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 |
| 6 | | | | | | -1 | -1 | -1 | 0 | 0 | -1 | -1 | 0 |
| 7 | | | | | | | 1 | 0 | 1 | 1 | 0 | 0 | 1 |
| 8 | | | | | | | | -1 | 1 | 1 | -1 | 0 | 1 |
| 9 | | | | | | | | | 1 | 1 | 0 | 1 | 1 |
| 10 | | | | | | | | | | 0 | -1 | -1 | 0 |
| 11 | | | | | | | | | | | -1 | -1 | 0 |
| 12 | | | | | | | | | | | | 1 | 1 |
| 13 | | | | | | | | | | | | | 0 |

4 The mean priority (MP) of a concept corresponds to the mean of the scores assigned to a concept with respect to the other concepts for all the participants. The value range of the mean priority is the interval [-100, +100]. This enables the distinction of the "positive" concepts and the "negative" concepts. For instance, if all participants provide the score +1 for a concept compared wit all the other concepts, the mean priority of this concept is +100.

The interparticipant consistency (C) of a concept corresponds to the mean of the standard deviations of all global scores. It is related to a consensus on the selection of subjective criteria. It conveys a degree of confidence in the MP with respect to the set of participants.

The mean priority deviation or stability of a concept (D) corresponds to the standard deviation of the mean priority with respect to the global scores of a concept. This parameter is useful to better characterize the priority concepts. In fact, the smaller the value of D, the more stable the group judgment

The global consensus (GC) expresses a global score of the group consensus on the investigated issue. This parameter is useful for comparing several groups.

Q1 *What did GEM bring to you?*
Q2 *Could you compare GEM to other knowledge elicitation methods (interviews, questionnaires, field studies, classical engineering analysis)?*
Q3 *What are the positive attributes of GEM (e.g., rapidity, creativity, efficiency, productivity, formal aspect, cooperative aspect)?*
Q4 *What are the negative attributes of GEM (e.g., brainwriting procedure, influence of the knowledge elicitation facilitator, non transparency of consensus calculations, shallow knowledge elicited)?*
Q5 *What do you find useful and unique about GEM?*
Q6 *In which kind of task would you recommend the use of GEM (e.g., requirement gathering, design, development, training, maintenance, marketing)?*
Q7 *Which phase of GEM would you like to see improved?*
*Q8 What should the knowledge elicitation facilitator know prior to the session (e.g., domain, task, problem, GEM)?*
Q9 *If this method was chosen by your organization, would you recommend training a knowledge elicitation facilitator to use it?*
Q10 *Would you reuse GEM by yourself as a knowledge elicitation facilitator?*
Q11 *Would you advertise this method to your colleagues?*
Q12 *Do you think that an extended computer-supported GEM would improve the current paper-supported GEM? Why?*

The answers to these questions revealed GEM's main advantages and flaws. We discuss these results in terms of four dimensions:

- Design as collaborative writing,
- Analytical and learning process,
- GEM as a semiformal mediating decision support, and
- The social construction of knowledge [1].

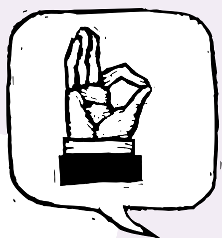
**Design as Collaborative Writing**
Writing has several advantages and drawbacks. Participants involved in a GEM session wrote short casual statements that usually included viewpoints that were stronger than if they had given the same information orally. Generated viewpoints are more or less specific episodes that can be subsequently endorsed or contradicted.

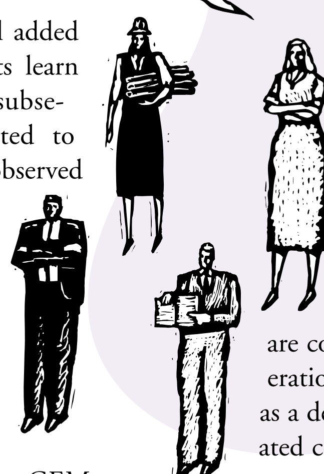
Sometimes problems occur because people cannot understand each other's handwriting. Participants may also interpret other participants' arguments incorrectly. However, this kind of flaw is eventually corrected during the reformulation phase (Phase 3). By trying to understand one another's terminology and conceptual analyses, participants incrementally construct and agree on an ontology of the domain. As the GEM progresses, participants start to write for others instead of writing for themselves by adapting their own language to what they have just read. This is called empathy [10]—that is, participants need to understand the needs and abilities of others if they want to communicate. The difference between talking and writing lies in the fact that participants have more time and opportunity to express their viewpoints in an undisturbed atmosphere. The properties described in this section are strongly supported by the answers of the participants to Questions 1, 4, and 7.

**Analytical and Learning Process**
During the viewpoint generation phase (Phase 2), by reading others' statements each participant allows his attention to be focused on the precise details of the argument. This read-and-write dialogue also involves both event-driven reactive behavior and reflective–deliberative behavior. In any case, reading leads to cognitive processes of interpretation, understanding, and reformulation, which are typical analytical processes. Positive reinforcement contributes to the validity of the viewpoint. A negative statement or refutation provides an alternative view. This alternative view is usually due to either a different background (the interpretation is different because the participant does not have the same knowledge as the originator of the initial viewpoint) or a different perspective (even if the current participant has the same knowledge as the viewpoint generator, she does not have the

same information). This result is supported by the answers to Questions 2 and 3. In particular, 76 percent of the answers to Question 3 support the cooperative aspect of GEM.

In this context, it is possible to talk about a logic of discovery. The indeterminate nature of the result of interactions among the participants as rational agents (ratification effect) creates a situation of investigation and confrontation of knowledge. By reading one another's statements and, in the later phases of GEM, by visualizing conceptual structures that the group generates, participants re-examine their own knowledge. The resulting effect is a crucial added value of GEM: participants learn from one another and subsequently are more motivated to accept the result. This was observed in each GEM session that we carried out and was confirmed by the answers to Question 5. In the beginning, most participants were skeptical about the use of GEM because they thought that they could not reach agreement. GEM not only provided a consensus derived according to normative parameters, but also an indisputable written agreement. All groups were very motivated and enthusiastic after each GEM session, as indicated in answers to Questions 10 and 11. Team building is a meeting activity that we found to be important in multiexpert knowledge acquisition; it further guarantees the social recognition of elicited knowledge (as supported by answers to Question 1).





## GEM as a Group Knowledge Design Process

Visual feedback enables participants to perceive explicitly the course of the knowledge design process. Participants see viewpoint generation as a sequential process, that is, they generate their own viewpoints and react to one another's viewpoints using the sequentiality of writing. Viewpoint generation is also a parallel process, because all participants generate viewpoints at the same time (reducing elicitation time). The difference between a regular meeting and this phase is that participants discuss viewpoints that are already explicitly expressed. The ontology of the domain that is explored has been cognitively constructed by reading and writing. This is significantly different from verbal expression of the same viewpoints; verbal expression is usually ephemeral.

GEM is based on social interactions mediated by an incrementally constructed shared memory and a measure of consensus (from answers to Questions 5 and 12). During the viewpoint generation phase, participants are collaboratively involved in a creative activity using written forms that are progressively filled in. During the reformulation phase, they cluster the viewpoints into concepts that are compared during the relationship generation phase (Phase 4). GEM can be seen as a decision-making process among generated concepts. Generated concepts are adequate solutions that are accepted by the participants. Consensus ratings are computed and can be used as mediating devices because they embody relative trends in the consensus to guide designers and evaluators in their final decision-making process.

Some of the answers to Question 4 pointed out that participants may introduce inconsistencies during the generation of relationships among concepts. This sort of problem usually results when participants do not have the same priority attributes in mind when they compare two concepts and then two others. Experience shows that the number of generated inconsistencies is usually small and that GEM sessions tend to converge toward satisfactory results, that is, results that are globally accepted by the participants during Phase 6 of GEM. Inconsistency checking can be extremely useful in measuring the validity of the results.

## Social Construction of Knowledge

As Winograd and Flores [14] noted, the rationalistic view of cognition is centered on the individual. In contrast, GEM allows several viewpoints on a given issue to be generated, cross-referenced, and refined. These viewpoints are incrementally reformulated and structured into concepts, which are socially constructed as written objects seen from different viewpoints. They can be used either as a basis for argument or taken for granted.

GEM differs from personal construct theory methods [6] that resulted in the incorporation of repertory grids as a major elicitation technique [2, 11]. Although GEM recognizes Kelly's dichotomy corollary that "[a] person's construction system is composed of a finite number of dichotomous constructs" [11], GEM tries to represent the hierarchical structure of Kelly's organization corollary that "[e]ach person characteristically evolves, for his convenience of anticipating events, a construction system embracing ordinal relationships between constructs." Unlike repertory grid methods, GEM does not attempt to characterize ordinal relationship types (internal attributes). Instead, each priority matrix represents a set of personal relationships (external attributes that need to be further determined by eliciting relevant attributes of each participant). In other words, instead of asking one participant to elicit explicit distinctions between concepts (internal attributes as semantic descriptors), GEM attempts to elicit individual ordinal relationships. These relationships may be typed by an examination of participants' profiles (external attributes as viewpoints or pragmatic context), which is usually made during Phase 6.

We elicited several criteria that were used for such a construction of knowledge during the reformulation phase:

- Simplicity. For comprehensibility, if the expression of a concept is too complex, it must be split into several simpler concepts. The simplicity of a concept is usually a function of the number of viewpoints that generated it.
- Interest. The interest of the participants in a concept is provided by the number of positive reinforcements of its various viewpoints. It may be considered also to be inversely proportional to the number of its strongly contradicted viewpoints.
- Robustness. The robustness of a concept is a measure of how much the participants support it. This, in turn, depends on the robustness of the initial viewpoints and backgrounds of the participants.
- Corroboration. By corroboration, we mean validation by exterior knowledge that is provided during reformulation discussions. We consider a concept corroborated if other concepts reinforce its constituting viewpoints.

GEM is both creative and normative, as supported by answers to Question 1. It is creative in that it enables the participants to generate new ideas that would not be possible to capture using a questionnaire. It is normative because of the consensus process generated. In particular, the normative aspect of GEM originates in the integration of a voting mechanism (generation of relationships between concepts), a concept clustering technique for reformulating viewpoints as concepts, a consensus formula based on priority matrices, and priority parameters that provide guidance in the elicitation and decision-making process. On the basis of the answers to Questions 8, 9, and 10 and our own experience, a good knowledge elicitation facilitator must have implemented GEM a few times in order to reasonably grasp the technique.

## Conclusion

This article presented the Group Elicitation Method (GEM), which takes into account various viewpoints of a group of potential users. It has been extensively used for user interface design and evaluation. GEM facilitates the gathering of people with diverse backgrounds. Compared with other methods, such as structured or prompted interviews, card sorting, or Twenty Questions [12], GEM integrates group meeting properties, analytical benefits of writing, and computer-supported decision-making guidance [8]. A GEM session typically takes no more than a full day.

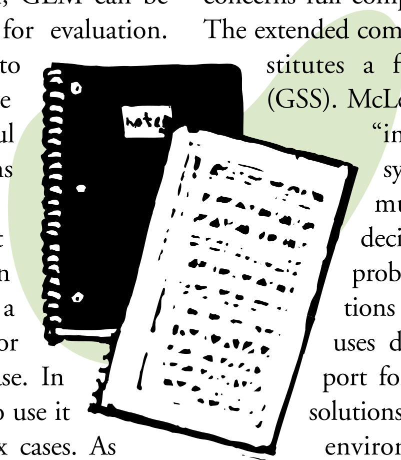
ACKNOWLEDGMENTS

Many people contributed to the current state of the GEM. Among them, I owe thanks to Catherine Logé, who helped me develop the first version of the GEM assistance programs; Rosel Schmengler. who strongly encouraged the use of GEM at Air France for an innovative experiment; Jean-Pierre Daniel. who helped me organize several GEM experiments at Aerospatiale on the assessment of the use of electronic mail and the development of a corporate memory; Jean-François Schmidt of Airbus Training, who provided us with a very interesting application on the redesign of an instructor workstation; the FANSTIC group that participated in the work that is illustrated here; and all the participants of various GEM experiments that were and will be carried out for the improvement the method. Alfred Attipoe, Markus Durstewitz, Martin Hollender, Rachel Israel, Laurent Karsenty, and Gene Lynch provided very useful feedback on the use of GEM. David Novick and Helen Wilson helped me to improve the quality of this article. Finally, I would like to thank Michael Muller who provided interesting comments on the final version of this article and persuaded me to publish it in *interactions* magazine.

This is a great advantage because results can be assessed directly, and expert end users are extremely difficult to convene for design or evaluation purposes.

As a decision support tool, GEM can be used for design as well as for evaluation. Although we have used GEM to design new systems, we have observed that it would be useful for evaluating existing systems and suggesting alternatives.

We have observed that GEM is a method that relies on the skills and the expertise of a knowledge elicitation facilitator during the reformulation phase. In other words, one must learn to use it before applying it to complex cases. As part of this process, a checklist is developed that states the issue to be investigated. Four types of operations are provided to guide the reformulation of viewpoints into concepts. GEM currently uses a minimal set of computer programs that help the knowledge elicitation facilitator to suggest directions in the knowledge design process. Charts and tables showing derived concepts using typical parameters such as mean priority, interparticipant consistency, stability of a concept, and global consensus provide relative trends and a basis for problem solving and decision making.




A question that arises from these sessions concerns full computerization of the method. The extended computer-supported GEM constitutes a full group-support system (GSS). McLeod [7] defines a GSS as an "interactive computer-based system that combines communication, computer and decision technologies to support problem formulation and solutions in group meetings." GEM uses decision technology to support formulation of problems and solutions in a brainwriting meeting environment. We have already developed a system that meets all these requirements. We found that industrial participants still prefer to write on paper using a pen and to have the knowledge elicitation facilitator write on a whiteboard and interact with them directly. The aim of our current research is to investigate new computer tools that will integrate these requirements, in particular the impression of the physical presence of the other participants and the facilitator.

### References


[1] Barrett, E. *Sociomedia: The Social Creation of Knowledge.* MIT Press, Cambridge, Massacusetts, 1992.

[2] Boose, J.H. Personal Construct Theory and the Transfer of Human Expertise. In *Proceedings of the 1984 American Association for Artificial Intelligence (AAAI-84) Conference.* American Association for Artificial Intelligence, California, 1984, pp. 27–33.

[3] Boy, G.A. *Intelligent Assistant System.* Academic Press, London, 1991.

[4] Boy, G.A. The Group Elicitation Method: An Introduction. In *Proceedings of European Knowledge Acquisition Workshop - EKAW '96,* Nottingham, United Kingdom. Lecture Notes in Computer Science series. Springer-Verlag, Berlin, 1996.

[5] Fisher, D.H. Knowledge acquisition via incremental conceptual clustering. *Machine Learning* 2 (1987), 139–172.

[6] Kelly, G.A. *The Psychology of Personal Constructs.* Norton, New York, 1955.

[7] McLeod, P.L. An assessment of experimental literature on electronic support of group work: Results of a meta-analysis. *Human–Computer Interaction* 7 (1992), 257–280.

[8] Neal, L. and Mantei, M. Computer-Supported Meeting Environments. *Tutorial Notes. Conference on Human Factors in Computing Systems, INTERCHI'93.* Amsterdam, 1993.

[9] Nielsen, J., Mack, R.B., Bergendorff, K.H., and Grischlowsky, N.L. Integrated Software Usage in the Professional Work Environment: Evidence from Questionnaires and Interviews. In *Proceedings, Conference on Computer– Human Interaction 1986 (CHI'86),* Mantei M. and Oberton P. (Eds.), ACM Press, New York, 1986, pp. 162–167.

[10] Norman, D.A. *Turn Signals Are Facial Expressions of Automobiles.* Addison-Wesley, Reading, Massacusetts, 1992.

[11] Shaw, L.G. and Gaines, B.R. Personal Construct Psychology Foundations for Knowledge Acquisition and Representation. Knowledge Acquisition for Knowledge-Base Systems. In *Proceedings of European Knowledge Acquisition Workshop - EKAW'93,* Toulouse, France, N. Aussenac, G. Boy et al. (Eds.). Springer-Verlag, Berlin, 1993, pp. 256–276.

[12] Tognazzini, B. *Tog on Interface.* Addison-Wesley, Reading, Massschusetts, 1992.

[13] Warfield, J.N. Societal Systems: Planning, Policy and Complexity. John Wiley and Sons, New York, 1971.

[14] Winograd, T. and Flores, F. *Understanding Computers and Cognition—A New Foundation for Design.* Addison-Wesley, Reading, Massachusetts, 1986.