

LRI Report No. 27

EVALUATING INFORMATION SYSTEMS

Peter G. Ossorio

Dan Popov

1981

This report was prepared for the Rome Air Development Center under Subcontract No. 9006-109-E03-12-12 with Booz-Allen and Hamilton, Inc.

LRI Report No. 27

EVALUATING INFORMATION SYSTEMS

Peter G. Ossorio

Dan Popov

1981

This report was prepared for the Rome Air Development Center under Subcontract No. 9006-109-E03-12-12 with Booz-Allen and Hamilton, Inc.

LRI Report No. 27

EVALUATING INFORMATION SYSTEMS

Peter G. Ossorio

Dan Popov

1981

© Copyright 1981
All rights reserved

Linguistic Research Institute
Boulder, Colorado

This report was prepared for the Rome Air Development Center under Subcontract No. 9006-109-E03-12-12 with Booz-Allen and Hamilton, Inc.

TABLE OF CONTENTS

Introduction	1
1.0 A Survey of the Human-Computer Interface Literature	4
2.0 A Conceptual Model for Evaluating Information Systems	24
3.0 Methodological Elaborations of the Conceptual Model.....	31
4.0 A Procedural Model for Evaluating Information Systems	43
5.0 An Examination of the METER System.....	52
6.0 Future Research and Development	66
References	82
Appendix A: State of the Art Evaluation	A-1
Appendix B: Material on Precaution Paradigm Simulation Paradigm, and Actor-Observer-Critic	B-1
Appendix C: Material on Parametric Analysis and Paradigm Case Formulation	C-1
Appendix D: The Logic of Evaluation	D-1
Appendix E: A Behavioral Approach to the Evaluation of Information Systems.....	E-1
Appendix F: Tradeoffs.....	F-1
Appendix G: Glossary	G-1

Introduction

In recent years the number and variety of information systems has increased greatly. This proliferation has created problems of optimizing human interaction with information systems and, correspondingly, problems of evaluating particular information systems or functional configurations within a general informational context. Such problems are particularly marked in situations where large scale, real time or near real time systems must be interfaced with human information processing. The USAF I&W system is a case in point.

The purpose of the present effort is to (a) survey current work on human-computer interface problems, (b) develop a general model for evaluating information systems, (c) examine a given information system in light of that model, and (d) suggest directions for further research and development.

The structure of the present report reflects the following conclusions based on a survey of the information processing literature.

- (a) Little attention has been given to the conceptual and methodological issues involved in either evaluation or interface problems. Rather, problems have been dealt with primarily and extensively at a technical level. Interface efforts have been directed toward either increasing the capabilities of the system or making the use of the system more convenient for Users. Presumably either accomplishment represents an advance in the state of the art.
- (b) The logic of evaluation does not appear to need revision in order to accommodate problems of evaluating information systems. However,

there is a good deal of confusion and misunderstanding in regard to the concepts and phenomena of evaluation.

(c) Genuine problems in the field of evaluating information systems appear to arise at conceptual, methodological, and technical levels. These problems reflect the following.

- (1) More than in many other fields, the field of information system usage is characterized by a lack of uniformity and consensus concerning criteria. The general methodological problems (and the associated practical complexities) associated with multiple criteria are therefore especially marked.
- (2) The criteria lie in the domain of human purposes, preferences, and activities.
- (3) We do not have the observational and manipulatory access to the criterion domain that classic evaluation requires.
- (4) Because we do not have the traditional observational/manipulatory access, we require an adequate conceptualization of the behavioral/social phenomena, an effective method of representing it, and a more sophisticated empirical methodology with effective technical implementation.
- (5) The resources for meeting these requirements are not readily available to practitioners in the information sciences.

As a result of these reflections we have made an effort to formulate a model for evaluation in the simplest and most direct fashion without regard to the convenience of any particular audience (another interface problem). The emphasis is on bringing together the necessary elements as clearly and coherently as possible. Unfortunately, this policy results in sometimes stating

what is perfectly obvious and sometimes making casual reference to systematic formulations which are relatively esoteric for many readers. In regard to the latter we have provided background material in a set of appendices which are frequently referenced in the text. The order of the appendices is designed to have some heuristic value.

The following are essential concepts which are likely to be unfamiliar or in need of clarification. Also indicated for each is the portion of the present report in which it receives its primary explication.

Actor-Observer-Critic	Appendix B
Configuration Description (structure-function description)	Appendix B
Descriptive Appraisal	Appendix D
Double Negative	Appendix D
Evaluation; appraisal	Appendix D
Paradigm Case Formulation	Appendix C
Paradigm Case	Appendix C
Parametric Analysis	Appendix C
Process Description	Appendix B
Precaution Paradigm	Appendix B
Simulation Paradigm	Appendix B
Task Analysis	Appendix B
Version; Version Class	Appendix E

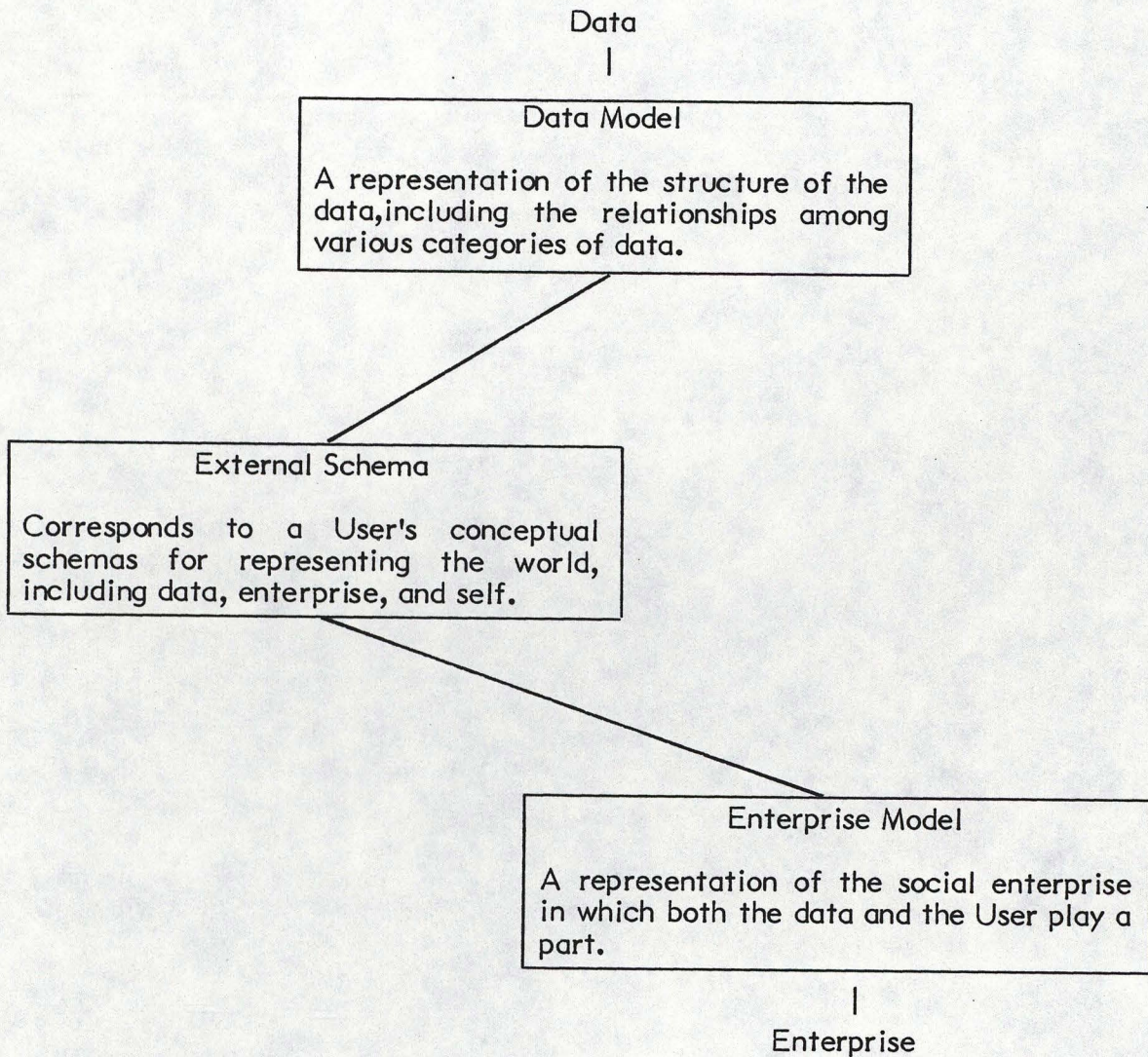
For convenience, a glossary of terms is included as Appendix G.

1.0 A Survey of the Human-Computer Interface Literature

The human-computer interface involves the following elements.

- (a) Data in a data base system is structured in a particular way, e.g., the tabular format of a relational data base. The systematic representation of this structure is commonly designated as a "Data Model" (see Figure 1).
- (b) The Data Model, together with the system software and hardware, imposes strong constraints on which information can be obtained from the information system (IS) and on how (by what User procedure) and in what form (graphic, verbal, etc.) any particular information can be obtained.
- (c) The information in the system is used by individual Users in the course of some set of human activities (the social system). The structure of these activities imposes strong constraints on which information is needed and when it is needed. Systematic representation of this structure is commonly designated as an "Enterprise Model."
- (d) The individual Users also have strong constraints on which information in what form is acceptably usable and on what procedures for information getting they are capable of engaging in with acceptable effectiveness. The systematic representation of such constraints is commonly designated as an "External Schema."
- (e) The constraints provided by the Users and by the activity structure are in general different, and often radically different, from those provided by the data structure. Correspondingly, the particular distinctions (concepts, entities, et cetera) which are to be found in an appropriate Data Model are in general different from those which appear in an appropriate Enterprise Model or External Schema.

Figure 1. Data and Enterprise



In general, the Data Model is different from and incommensurate with the Enterprise Model, since they have radically different subject matters. The User's conceptual schema comes into play in the attempt to map Data Model descriptions onto Enterprise Model descriptions. Such mapping would be facilitated if the Data Model and the Enterprise Model employed a common conceptual system.

- (f) The differences in the structure and content of the Data Model, the External Schema, and the Enterprise Model present two kinds of problems. (1) The first is how to transmit information effectively between information system and Users, given these differences. (2) The second is how to reduce these differences. These are the human-computer interface problems.

The psychology of designing computer interfaces focuses on the reaction of the User, and in particular the novice User to the man-machine system. In this context is found the contrast between User expectations with machine capabilities (Sackman, 1970; Martin, 1973; Schneiderman, 1980). In addition to the problems resulting from extreme assumptions on the part of the User as well as inaccurate assumptions there is often the problem of perceived role reversal. This reversal of subject and object leaves the User feeling that he is being molded by the system rather than accessing a useful tool. This perceived loss of control can result in a number of unfortunate results. The User can come to see the system as unresponsive to the needs for which it was created, and as being unpredictable and unrealistically fragile (Elam, 1980). Differentiating between Users and personalizing interactions with Users are methods that have been adopted by system designers to make the system appear more responsive to the User.

Setting out to design a User-friendly interface is an art form because of the lack of much systematic research in the area (Hackathorn & Zeiger, 1980). The computer in conjunction with the User ought to perform a number of functions in such a system. The User-friendly man-machine system ought to provide for computation, specification of needed content, search procedures, inference on the basis of retrieved data to assist in drawing

conclusions, assimilation of the conclusions as new information, explanation of this understanding in terms comprehensible to the Users and others. Rarely does a system designer attend to more than the first two considerations (Keen, 1979). Systems designed with this criteria in mind focus on the initiative element of the system, whether the initiative is provided by the User, the machine, or is mixed between the two. At the interface this is manifested in the dialogue mode selected for the input/output. The most common dialogue modes are: question and answer; form filling; menu selection; function key with command language; User initiated command language; query languages; natural languages; and interactive graphics (Atwood & Kirschbaum, 1979). Each of these requires and/or permits a number of types of Users in order for the system to remain User-friendly.

Research into these areas has taken a number of divergent directions ranging from display design to the underlying conceptual framework of the data base structure. All approaches have a common element of attempting to provide a comfortable, free and productive dialogue between man and machine. The goal of such a dialogue is the elicitation of shared understanding between man and machine (Shaw, 1979). Whatever mode is selected to promote such a dialogue must adhere to the four principles of expectation, implication, experimentation, and motivation (Jones, 1978), or some other approach to dialogue that is based on how human beings engage in a meaningful dialogue (Fitter, 1979). These factors hold true even when the system is the initiator of the dialogue using a display to initialize the interaction.

In terms of designing output displays for the purpose of transmitting technical data to enable a User to better interact with a computer, a number of problems have been identified. Limitations on display size result in data being

highly encoded and difficult to interpret, resulting in problems of learning and memorizing the display. Other issues include location on the CRT, contents of display (alphanumeric or graphic), and conversational vs. formatted displays. These problems are magnified in real time situations where speed and its associated human stress become an issue. While the design considerations are fairly well understood in addressing these issues, the human needs and limitations critical in system design are much less well understood (Rosenthal, 1980). The number of symbols displayed as well as the color in which they are displayed can make a difference in the interchange (Carter, 1977; Zimmermann, 1976), as can the sequence in which they are displayed (Badler, 1976). In order to begin to approach an understanding of the complexity of these issues and the role they play, a much more systematic operational analysis than has previously been conducted is called for (Lees, 1977).

Even more innovative programs allow material to be displayed in a tactile mode that allows the User to "feel" a three-dimensional display using a sticklike mechanism controlled by a servomotor system (Bell Labs, 1975), or hear the output of a calculator or terminal designed for the blind (Savoie, et al., 1976; Kutsch, 1977). The impact of auditory output is being studied for other groups of Users as well as the qualities of this output needed to transmit meaningful information (Witten, 1977). Another form of output or display, that of hard copy, has a future dependent upon advances in the field that will enable it to keep pace with these technological alternatives (Wieselman, 1977). While the displays may 'initiate the dialogue,' inputs are equally critical to maintain the interactive element. When a User accesses a computer, one of the first things confronted is the disparity between most interactive computer languages and natural language in both vocabulary and grammatical rules. In order to

alleviate this disparity two methods have been suggested. These methods, originating from linguistic theory, involve the use of a dependency grammar and the use of interpretive grammar (Rohlf, 1980). Both these methods describe the computer language in a way that enhances understanding of the needs of the User early in the design process so that inappropriate languages might be avoided. Rather than selecting the interactive language, programs designed to rapidly analyze speech are also being developed (Witten & Griffith, 1979). These approaches appear promising in both speed and cost effectiveness.

The primary trend today is not to select the language in support of a particular task but rather to develop higher level interactive languages that more closely approximate conversational English. In these fields reliance on syntactic and semantic processors are used to process the language (Guida & Somalvico, 1980). The processing of natural language has been developed in the field of artificial intelligence to allow the naive User to extract information from computerized data without the need for a technical intermediary, be it a programmer or a higher level query language (Harris, 1978). Techniques in the field of using natural language as a query language were pioneered by Winograd (1972) and Woods et al. (1972). In all the developed techniques there is a distinction between using natural language to query a data base and to use natural language to generate programs within the data base. A problem in using natural language as a query system is the ambiguity in the language itself. There are times, however, when this ambiguity can be an asset as well as a liability (Thomas, 1978). METER (Brown et al., 1979) uses a number of pragmatic short cuts such as deleting suffixes, etc., in order to simplify language processing while retaining the ability to use unformatted English to form a search query. Not all researchers see natural English as a viable basis for an interactive

language, and advocate the development of a general purpose query language whose syntax and grammar approximate that of English (Fitter, 1979).

When a natural language is envisioned as a means of input/output, a logical extension of this is the development of a voice recognition program. In many circles this adjunct to natural language interactive systems is seen as one of the ultimate tools (Herscher & Cox, 1975). In the field of speech recognition the notion of knowledge is first introduced and its uses to generate and verify hypotheses are explored (Reddy, 1976). Much of this emphasis on voice recognition and the use of natural language as a query language are due to its perception as the most "natural" mode of man-machine dialogue. Another way that the man and machine can interact in a productive interchange is via computer graphics.

More and more, graphic interaction is being used as a mode of interaction between man and the machine (Herot, 1980). These approaches enable the User to have more flexible access to the contents of the data base. It has a particular attractiveness to naive Users who can access data in the system in a mode that encourages browsing and thus added familiarity with the data base architecture. This trend toward more consideration to the human factors involved in graphic displays has been a phenomenon of the last ten years (Spence, 1972; Humcke & Kent, 1977). Since the first major application of interactive computer graphics in 1963 (Sutherland & Johnson, 1963), the focus of research and development has been in the area of software and hardware. Recent research has focused on defining a man-machine environment in which individuals who are comfortable with tasks that need to be performed are selected and assigned tasks commensurate with their training. Additional study has been done on environmental layout, lighting, and hardware configuration

(Welt, 1975; Aronson, 1976; Topmiller & Aume, 1978). In addition to these traditional human factor studies, studies on shift work operations and decision making have proven to be relevant as well (Cooley, 1976; Maver, 1976). Much less work has been done on non-traditional human factor issues, particularly those that deal with the general non-specialized User.

Hardware advances have been responsible for many of the advances in the use of graphics as an input/output source. Light emitters such as CRTs and light controllers such as liquid crystal displays have provided control of display, brilliance, etc. and have made it possible to configure a hardware system tailored to almost any application imaginable (Grover, 1977). In addition the problem of image resolution is being addressed (Giese et al., 1978). Indeed, the application of computers to the system design process itself has been somewhat of a revolution. This area of computer-aided design (CAD) has allowed rapid development in both hardware and software design (Llewelyn, 1977). Indeed this same computer-aided design process has been applied to the human factors (Spence, 1976). This blend of human factors studies and CAD has delimited many of the parameters of interest in the area of interactive graphics. Such elements as system functions, the perception of pattern, art, computer response time, User reaction to delay, uniqueness of display are all design considerations for system designers taking human factors into account (Spence & Apperby, 1974; Negroponte, 1970; Goddard et al., 1971; Newman, 1968). These considerations culminate in the general area of command dialogue. Here the criterion becomes the ease of use of the system defended in terms of the User (Martin, 1973; Foley & Wallace, 1974; Carlye et al., 1974).

One of the primary problems associated with graphic displays is the need for fast update of the display, particularly in interactive situations when

the displays are of high resolution (at least 500 x 700 pixels). One of the approaches to the problem is a distributed architecture that can transmit a pixel in 500 ns (Rothsberger & Rakotomalala, 1980). Other programs have depended upon prompts at critical times in performing complex graphic operations (Paisner, 1979). Still others have approached this issue by improving upon the speed of the internal methods used to generate the graphics (Stocker, 1979). While many approaches have been generated to address the speed of the interaction between the User and computer in graphic displays (McCann, 1978), progress in this field has been limited by the lack of a comprehensive understanding of User needs and limitations.

Another facet of graphic displays that leads to potential interface problems is the need for the User to specify the curves or data to be displayed, often in a procedural language far removed from that process typically used to generate a graph. One approach to this problem is the development of an interactive series of questions and User answers, the results of which are used by the program to generate journal quality plots (Drake, 1980). While such customizing options are becoming more common and useful, basic plotting packages are still the major method of generating graphics (Maller, 1980). Common packages in the field today involve direct transfer of graphic material, such as the low cost system that makes use of an overlay tablet (Corbett, 1976). In general the field of computer graphic input/output structures has been developed in isolation from other I/O systems and from the general issues of man-machine and communication philosophy. This is readily seen in the number of general, introductory texts and reviews being written in the graphics field (Encarnacao & Oldenbourg, 1975; Elliot-Green & Parslow, 1971).

Other innovations in this area involve data structure design (Cardenas & Seely, 1975), plasma displays (Jarvis, 1975), color displays (Jones, 1974), and a wide range of terminal-based displays that put an interactive graphics capability within the reach of almost any system User (Machover, 1974; Grover, 1977; Potts, 1975; and Reingold, 1974).

The input/output function is not, however, the only component involved in the man-machine dialogue. The data base architecture itself plays a major role in the dialogue. Recall earlier the distinction was made between using natural language for querying the data base and for generating programs to manage the data base itself. It is a goal of data base design to reflect the reasoning process of the human User within the computer itself (Hayes, 1978). Within the architecture of the data base itself, the relational structure of most large scale information systems attempts to replicate the conceptual reasoning process used by a human User (Schank, 1975).

The design of data bases introduces a wide range of interface issues, particularly when the human User is seen as an integral part of man-machine systems (Zeiger & Hackathorn, 1980). The perennial problem emerges of who will the Users be. The answer to this question often is critical to data base design. Record based data models, even those including sophisticated associative features, have some significant limitations (Kent, 1979). As in the input/output interface, graph-based models for data base design are beginning to emerge from the field of artificial intelligence (Winston, 1977; Nilsson, 1980). Programs such as FRAMES (Minsky, 1974; Bobrow & Winograd, 1977), SCRIPTS (Schank & Abelson, 1977), and the work on semantic networks (Quillian, 1968; Rumelhart & Norman, 1975) show a significant potential for providing an alternative to current record based systems.

The data bases involved in artificial intelligence approaches concern themselves not with record based storage or the relations among lists of data but rather with the engineering of knowledge (Fergenzaum, 1978).

The distinction between data, information and knowledge becomes critical here. Whereas relational file based systems involve the storage, retrieval, and processing of data, artificial intelligence systems involve the representation of worlds (Feigenbaum & Feldman, 1963). Here is where the bridges are to be made between the language of the query and the information that is contained therein (Gregory, 1968). It is the relationship between artificial intelligence, language, and knowledge that bodes well for the systematic representation of real world systems in the future (Goldstein & Papert, 1977; Fergenzaum, 1974; Ossorio, 1971, 1978). This is the promise of the future. Today artificial intelligence applications are extending natural language queries into the realm of data base management (Thompson & Thompson, 1978). The query systems of EQS (Martin, 1978) and METER (Brown, et al., 1980) use a blend of control language and the natural language query to access and retrieve data, but a full scale application of real world representation awaits a much more systematic model of knowledge and the human thought process that represents the real world than is currently extant (Vallee, 1976). The vast majority of data bases currently operational are record based systems, many with relational structures accessed not through a natural language but rather by a machine-based higher level query language.

One of the most common forms adopted to meet the needs for man-machine communication is that of the menu selection system. One of the approaches developed to explore the limits and potential benefits of this approach is ZOG (Robertson, et al., 1979). This approach has been tested in a

variety of task domains and seeks to explore rather than promote this particular communication philosophy. This approach, coupled with the alternatives of question and answer and form filling, are the three primary modes of interaction utilized at the present time. Each has advantages and disadvantages, with the advantages being maximized when the mode of interaction is a "familiar situation" for the User (Davies & Yates, 1979).

Unique problems arise when changes are made to existing data bases or a variety of diverse systems are being integrated into a single operating system. These design considerations are being systematically explored in a number of settings (Logicon, Inc., 1979). The issues here extend beyond the man-machine interface and into the realm of interfacing two independent systems in such a way that information can be exchanged regardless of the language used in the management of the data base. To do this in a convenient and productive fashion it is necessary, if not mandatory, to separate such a management system into two separate but interrelated models, that of an enterprise model and a data model (Levin & Schneider, n.d.). The enterprise model identifies the entities and relationships pertinent to the enterprise while the data model identifies the names and associations among names that comprise a data base. Rather than seeking a single integrated system, the data model is then mapped onto the enterprise model. The advantage of this approach is that the structure of the relational data base need not limit the input/output configuration of the interface (Schneider, 1978).

If the interface problem is approached in that fashion, then a means of mapping one model upon another is called for. This mapping founded on a relational model with a separate and distinct accessing or enterprise model would enable a User to have a separate view of the data represented and map

queries relevant to that view onto an optimized accessing program (Codd, 1970; Senko et al., 1972; Schneider, 1976; Ghosh & Senko, 1974; Codd, 1971). In order to implement such a program a higher level data base model is required capable of accessing at a number of levels of abstraction. Such a model is DIAM (Levin, 1978; Senko, n.d.; Senko, n.d.). This model has utility in separating out these two levels and mapping one on to another, and it makes it possible to develop a data independent model of data bases that can be used to simulate as number of data bases and serve as a test bed for experimentation (Schneider & Connolly, n.d.). A summary diagram of DIAM is shown below in Figure 2.

It is the promise of such testing and the dichotomous view of the enterprise and the data models that points to the possibility of developing a view of an information system as a system to integrate the User and the computer, each component of which is given equal weight without an attempt to either force the User into the machine mode or to represent in the machine the User per se (Schneider, 1976). Up to this point in time integration has not been as much the goal as comparing the utility of one paradigm to another (Van Gigch, 1979) or one approach to relational data base development to another (Brosey, 1978). Query languages have been compared to one another (Greenblatt & Waxman, 1978) and data models compared holding the interface constant (Lochousky & Tschritzis, 1977; Lochovsky, 1978), but experimentation on an information system awaits a model of such a system that is neither user nor machine dependent.

The enterprises, themselves, place limitations on the information systems and raise interface issues independent of both Users and machine. An information system designed as a decision aid or as a problem solving tool may

Figure 2. Summary of the Data Independent Accessing Model (DIAM)

DIAM LEVEL	DESCRIPTIVE LANGUAGE	MANIPULATIVE LANGUAGE	DOWNWARD MAPPING
ENTERPRISE	REAL WORLD ENTITIES AND RELATIONSHIPS	REQUESTS FOR DATA ABOUT REAL WORLD ENTITIES AND RELATIONSHIPS	SELECT FROM DATA RELATIONS THOSE THAT ARE MOST RELEVANT TO INFORMATION REQUESTED
DATA	STORED DATA DOMAINS AND RELATIONS AMONG THEM	ALGEBRA OR CALCULUS OF DATA RELATIONS	SELECT FROM ACCESS PATHS THOSE THAT ARE MOST RELEVANT TO NECESSARY DATA RELATIONS
ACCESS PATH	DIRECTED GRAPHS PRESENTING THE RELATIONS THAT ARE PHYSICALLY IMPLEMENTED	PROCEDURAL NAVIGATION FOR DIRECTED GRAPHS	DECODE ACCESS PATHS INTO COMPONENT SERIAL BLOCKS
ENCODING	BASIC ENCODING UNITS THAT DETERMINE HOW IMPLEMENTED RELATIONS ARE REPRESENTED	CONTIGUOUS DATA GROUP LANGUAGE FOR RESOLVING DISCONTIGUITY INTO SERIAL BLOCKS	COMPOSE SERIAL BLOCKS INTO ADDRESSED STORAGE ACCESS
FRAME	CORRELATION BETWEEN SERIAL DATA AND SERIAL STORES	ADDRESS SPACE LANGUAGE FOR ACCESSING DATA STORES	TRANSLATE STORAGE ADDRESSES INTO DEVICE INSTRUCTIONS
DEVICE	ASSIGNMENT OF SERIAL ADDRESS SPACES TO MEDIA DEVICES	COMMANDS TO ACTIVATE DEVICES AND CHANNELS	

well call for a design far different from a system designed for a library function or that designed as a real-time simulation.

A major application of computers in this light is as a decision aid. One of the limitations of this use of the computer is the ability of the User to interact with only one computer program at a time. This limits the User to only one source of information at a time and to take advantage of one computer process at a time. One approach to this problem is that of the Windows Management System (Oppenheim et al., 1980) which allows a User to access a number of processes at a time from a single terminal. When this issue is addressed the various roles that both the User and the computer play at various times impact on systems design (Rouse, 1977). Here the question of overlapping responsibility of each becomes paramount, as does the User's ability to attend to and hold the information being displayed (Enstrom, 1977; Kleinman, 1977). Methods used for problem solving (Jandy, 1977) and the development of plans of action appropriate to the task at hand based on information generated at a higher level of abstraction impact directly on the effectiveness of such systems (Shaket, 1977). Here the design of a system to be used for problem solving or as decision aid can call for a dialogue far different from a system designed for a mere exchange of data where the social systems equivalent of master-slave or consultative relationships would be better descriptions of the interchange between man and machine and where pursuit of the correct information may be as critical as its accessibility (Ruer et al., 1978). Even in these systems, the choice often lies in selecting a rule based on a structured approach to decision making vs. a logical-reason-based artificial intelligence system (Anderson, 1978).

Another enterprise that presents some unique challenges and issues is that of simulating some aspects of the real world. The hardware and software for such systems are becoming so available that simulation of physical systems (i.e., weapon systems, etc.) are becoming commonplace (Walsh, 1976). Uses of simulation in training settings have demonstrated the validity of such models (Leonard, 1975) as well as in operational systems (Vaccarino, 1978). Less well validated and yet perhaps most critical is the simulation of human and social systems. Systems in these areas have been developed most often to meet military decision needs (Ashton, 1977). Perhaps the greatest challenge for this type of modeling is to take on the task of modeling the information system itself. The goal of this enterprise would be to develop models of sufficient complexity and rigor to allow their use in optimizing the information system itself. Some work has already been done in this area. It is the issue of optimizing the man-machine interaction that determines the efficiency of the system (Romov & Venda, 1977). The focus of such models may be on the operator and User strategies for optimizing the process (Laughery, 1977) or on the design factors of the machine systems that work to optimize the interaction (Williges, 1977). In either case the goal is the same, to get a grasp on the complex components that make up the information system and the even more complex interactions between these components which determine both the efficiency and effectiveness of the man-machine interface. What is obvious from the literature is that a system model capable of addressing this complex task is not currently available.

What is available at the present time is a number of programs designed to integrate the User with existing or planned systems. Sophisticated systems of this sort rely upon the computer to tutor a potential User in the use

of the machine. These systems draw heavily on the field of computer assisted instruction for their theory and methodology (Richardson et al., 1980). However, even in this fairly well established field there is little or no standardization of approaches, models, hardware, and software to include the languages most appropriate to CAI applications (Chambers & Sprecher, 1980). Design considerations for self-teaching computer systems are using a language familiar to the Users with a limited grammar and vocabulary (Kennedy, 1974), being sensitive to User attitudes during training (Gedye, 1972) and continual monitoring of the User's learning performance during the training process (Gaines & Facey, 1975). When these considerations are kept in mind during system design, training the naive User is a fairly efficient task requiring a fairly small number of sessions (Kennedy, 1975).

It is perhaps only in this area of computer assisted instruction that traditional psychological considerations play a significant role in the literature. In this area, feedback (Prother et al., 1972), anxiety (Denny, 1966; Dunn, 1968), and the human learning literature play a major role in system design. Herein lies the challenge for man-machine system design to integrate the human component into a system without compromising the User or the system (Nickerson, 1969).

Whenever an input operation occurs, there is the danger of a human error on input. These common input errors have been classified and approaches to error reduction discussed (Mace et al., 1979). Preparing data for input in a particular way is another method for reducing errors (G.A.O., 1978). These methods are becoming increasingly more viable as keypunching and the punched card are being replaced by more advanced systems. Each system (menu, voice recognition, and keyboard entry) has unique advantages in error reduction of

various sorts (Welch, 1977). No clear-cut advantage of one system over another has been demonstrated. In addition, although the reliability of machines continues to rise with every hardware improvement, the reliability of the User stays constant and remains a concern for both operators and systems designers (Swain & Guttman, 1975). Even if this issue were managed there would remain the inevitable sources of indeterminacy inherent in systems of this sort (Meister, 1975). Thus the challenge to both the researcher, designer, and User of man-machine systems continues. As more is understood about how systems behave, the focus of effort may be expected to shift from how to make such systems function well, to how to prevent a well functioning system from going wrong in one of the ways in which it can.

In retrospect, the interface issue can be (and quite often is) seen as the province of display designers, input/output architects, data base management systems, conceptual theoreticians, query language designers, human factors engineers, input/output hardware and software systems, and social scientists. It is an interesting idea to take the User into account when systems are designed, if one can define who the User is. If one is unsure or unable to specify the User in terms that make sense within the system as a whole, it is equally attractive to design information systems to maximize the capability of the machine (keeping the User-friendly guidelines in mind, of course) and to integrate the User into the system either through a selection and/or training process. This fragmentation of experimentation, research, and system design considerations continues to make the design of an information system that maximizes the integration of the User and the computer by optimizing the man-machine interface more of an art than a science. Reducing the issue to one of "human factors" in what is otherwise inherently a machine-oriented

system fails to acknowledge that the purposes or enterprises for which the information system were created in the first place are human enterprises.

In the review of the literature it quickly became apparent that a major need in the field is for a balanced view in which each aspect or component (User, display, hardware, language, etc.) receives appropriate acknowledgement and representation rather than a preemptive, myopic, or reductive representation. Further, it was evident that progress in this respect has been limited by the absence of a comprehensive systematic model for representing the information system as such functioning within the social context for which it was designed. Until such a model is developed, validated, and used systematically to explore the impact of various component changes on the functioning of the man and machine system as a whole (the social enterprise), User-friendly design or effective functioning of man-machine information systems are likely to remain more of an art than a science.

The problem appears to have a familiar basis, i.e., that of the fragmentation of science and technology into more and more specialized and less and less interconnected fields. Once we define data bases, query languages, displays, human factors, social systems, etc. as separate subject matters, we evolve for each a separate model or set of models, each having its own distinctive concepts and relationships and having no conceptual connection to any other domain. We have not been able to put Humpty together again.

There is, however, one sort of conceptual framework, and quite possibly it is the only one, in which data bases, query languages, social systems, et alia are interconnected without prejudice to the technical developments within each. This is the framework from which they emerged to begin with, the framework in which we conceptualize the ordinary real world as such, i.e., a

domain which comprises people, social systems, machines of various sorts, languages, planets, countries, natural and artificial terrain features, oceans, information, animals, plants, theories, satellites, armies, and so on. A conceptual-notational system which provided systematic representation of this domain could be expected to have the necessary formal characteristics for "representing the information system as such functioning within the social context for which it was designed."

Such a conceptual-notational system has been developed within the emerging field of Descriptive Psychology (Ossorio, 1978). The remainder of the present report is an application of Descriptive-Psychology concepts and methodology to the problem of validating information systems or features of such systems. It addresses directly the problem of exploring the impact of various component changes on the functioning of the man and machine system as a whole. It also delineates rationales and procedures for trading off various degrees of evaluational rigor for practical constraints and cost effectiveness.

2.0 A Conceptual Model for Evaluating Information Systems

A conceptual model for the evaluation of information systems is readily derived from the general concept of evaluation elaborated in Appendix D. The elements of this model are given below and a procedural illustration is shown in Figure 4 at the end of this section.

- 2.1 The information system (IS) which is to be evaluated is embedded in a social system (SS). The functioning of the IS is part of the functioning of the SS and contributes to the functioning of the SS. It is this contribution which is the primary thing to be evaluated about the IS.

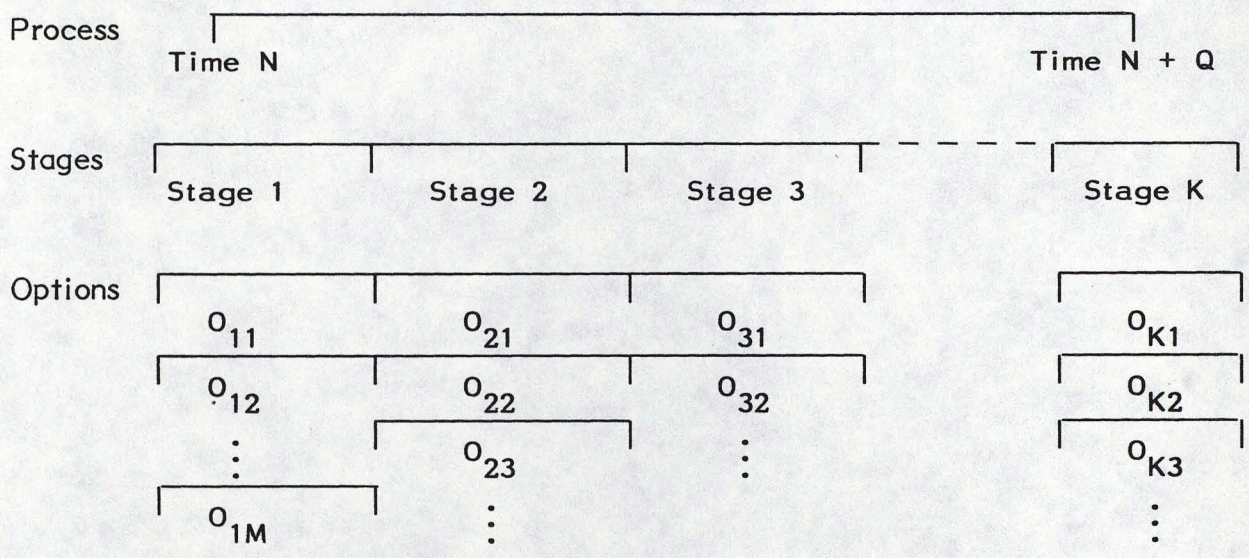
Figure 3 makes use of some of the resources of Process Representation (see Appendix B) to show some of the relevant relationships in visual form. It illustrates, for example, a systematic way of representing that the social system functions differently depending on certain differences in the IS functioning, or on its presence or absence. These dependencies, or Contingencies, correspond to "the contribution of the IS to Social System functioning." It is a general feature of Real World Representation that it systematically relates objects, processes, events, and states of affairs and that it systematically represents part-whole phenomena within and across these categories.

The social system is whatever structured set of human relationships and activities calls for the operation of the information system. For example, it may be the entire I & W network, the activities of a particular installation, a particular project, and so forth. Relativity to some context is an irreducible feature of evaluation (Appendix D). The social system is the context which provides an actual, direct criterion

Figure 3. Information System in Context

Process Name: Functional Cycle of Social System

Process Description:



- Elements in stage/options
- E₁ = User₁
 - *E₂ = IS
 - E₃ = ship
 - E₄ = User₂
 - E₅
 - .
 - .
 - .
 - E_R

This figure is an incomplete process representation. It represents the gross functioning of the social system as a sequence (stages) of possible happenings (Options, which are themselves processes) involving a given set of Elements at each stage. An information system would appear as one such Element, as would a User. The technical representation of the IS functioning would appear in a subprocess description (e.g., a component of O₁₂) in which the Elements were, e.g., the components of the Data Model. The contribution of the IS would be given by Contingency specifications connecting IS process options to SS process options. The sequence O₁₂, O₂₃, O₃₁, . . . O_{K3} is a Version, or possible history, of the Social System functional cycle. Process representation is not flow-charting.

for evaluation as contrasted with various indicators, which are indirect, or subsidiary 'criteria,' which are more or less arbitrary.

- 2.2 "The functioning of the social system," which is what the IS contributes to, may be represented by the set of possible histories of the SS during a functional cycle.

"A functional cycle" refers to a historical unit in the history of the social system such that it makes sense to speak of repetitions of this unit within that history. The unit may be a "natural" one such as a type of military mission or it may be a conventional one such as "one week of operations."

- 2.3 Each of these possible histories is designated as a "Version" (see Appendix E) of the functioning of the social system. The set of Versions is the set of possible histories.

The set of Versions can be expected to be too large and heterogeneous to be characterized and evaluated. If so, the members of the set must be classifiable into a manageable (e.g., 5-15) number of Version Classes (see Appendix E).

Version classes must be capable of being directly evaluated. That is, given a specification of the Version Class and the question "How would it be if things went that way?" it must be possible for knowledgeable or interested judges to answer, e.g., "Acceptable," or "Unsatisfactory," or "Disastrous," and so on. A limiting case here would be the simple dichotomizing of all the possibilities into two Version Classes, i.e., "Successful" and "Unsuccessful" or any equivalent thereof.

Effective classification of Versions into Version Classes may be accomplished by means of Paradigm Case Formulations (see Appendix C) with selected scenarios as Paradigm Cases (see Appendix E).

2.4 What is to be evaluated is the functioning of the IS.

"The functioning of the IS" is represented by the set of its possible histories during a functional cycle which is defined by reference to the social system functional cycle (2.2, above). Each Version of social system functioning includes a portion which consists of the IS functioning. For present purposes, the relevant functional cycle of the IS is its participation in the social system functional cycle.

If necessary IS Versions can be classified into Version Classes using the social system scenarios to define Paradigm Cases or using IS-specific scenarios.

2.5 Any element (e.g., the IS) which contributes causally or in a part-whole way to the functioning of the social system can be assigned an instrumental value so long as the functioning of the social system can be assigned value.

2.5.1 If the overall social system functioning is assigned a numerical or categorical (e.g., "acceptable") value, the components (including the IS) may be assigned the corresponding instrumental value. (This is part of the logic of "Don't argue with success!")

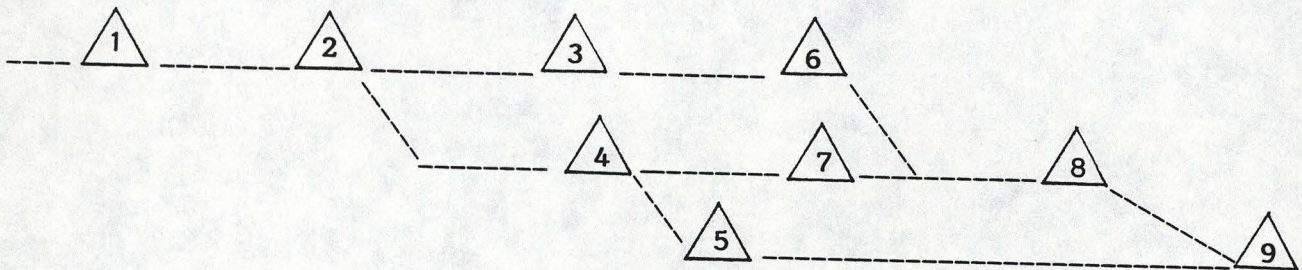
2.5.2 In a comparative evaluation the contribution of a component is compared with the contribution of an alternative "same" component. (For example, two different information systems may be compared as alternative components of the social system.) The comparative evaluation is accomplished by evaluating each alter-

native separately as described above and comparing the quantitative or categorical instrumental value characterizations.

- 2.5.3 The point of evaluating the IS across the possible histories (excluding the IS Version components) of the social system is to provide an assessment which grants the other sources of variability in the social system functioning. Conceptually, this is the equivalent of assessing the contribution of the IS under the condition "other things being equal."
 - 2.5.4 Any significant modifications of the IS (e.g., a change of software, addition of graphic or interactive facilities, etc.) can be evaluated by considering the modified and unmodified forms of the IS as alternative components.
 - 2.5.5 So long as individual Version Classes of the social system can be directly evaluated, in the same way that the overall functioning of the social system can be directly evaluated, the IS can be assigned an instrumental value relative to each such Version Class. So also can modifications of the IS. The latter procedure would be involved in tradeoff analyses.
 - 2.5.6 Conceptually, it would make sense to compare social system functioning with and without the IS. However, ordinarily such a project would merely confirm that the IS in some form is an essential part of the social system, and that would normally not be in question.
- 2.6 Formally, the assignment of numbers to value categories can always be done, but it may be more or less arbitrary and therefore useless or misleading in a practical sense. Assigning numbers to value categories may be empirically meaningful only within a limited range of IS exem-

plars or only relative to a given criterion (e.g., average level of function, catastrophe prevention, etc.). If a set of comparative evaluations can be quantitatively rationalized so that information systems with higher evaluation indices are always (practically speaking) superior to those having lower evaluation indices, then, within whatever range limitations are applicable, the evaluation index will be meaningful as an "absolute" (i.e., only implicitly comparative) index.

Figure 4. A Procedural Illustration of the Conceptual Model



1. Construct a Process Representation of Social System functioning, with the IS being an Element and its functioning being a subprocess
2. Define Paradigmatic scenarios for social system histories
3. Use scenarios to define Version Classes V_a, V_b, \dots, V_n for social system histories
4. Define Version Classes V_1 and V_2 for the IS functioning
5. Specify K , the feature of the IS which distinguishes V_1 from V_2
6. Evaluate V_a, V_b, \dots, V_n by direct judgment and generate the corresponding evaluative indices E_a, \dots, E_n
7. Specify the set, S_1 , of Version Classes among V_a, \dots, V_n which are contingent on V_1 and the corresponding set, S_2 , of Version Classes which are contingent on V_2 . (Frequency distributions may be used instead of simple sets.)
8. Specify a function (e.g., averaging) for combining evaluative indices E_i into a single evaluative index
9. Apply this function to the sets S_1 and S_2 to arrive at E_1 and E_2 .

The difference $E_1 - E_2$ is the value of K , i.e., the value of its contribution to social system functioning

3.0 Methodological Elaborations of the Conceptual Model

It is possible to give a structure-function representation of the social system and its functioning (Appendix B, process description; Ossorio, 1978; Jeffrey and Putman, 1981). In such a representation, particular abstract individuals or groups will have particular roles, or parts, to play.

This kind of representation corresponds to an implicit task analysis with respect to social system functioning. That is, conceptually we may start with the idea of that social system functioning as being an abstract state of affairs to be achieved and ask "How?" or "What would qualify as that?" The answer would take a form exemplified by "By having X_i , Y_i , and Q_i happen, and further, by having one individual, P, do X_i , having another set of individuals, R_i , do Y_i , and having another individual, Z, do Q_i by first doing G_1 then either G_2 or G_3 and if . . . doing G_k ."

A structure-function description will refer to abstract individuals because on different occasions a given part may be played by different historical individuals. On a given occasion a historical individual is empirically identical to an abstract individual who is defined by his part in the process in question.

3.1 A task analysis of this sort provides a new basis for evaluating the performance of historical individuals. A given individual may be evaluated in regard to how well the individual plays the assigned part. If the assigned part is "doing Q_i " or "being Q," there will be standards in regard to what qualifies as doing Q_i (or being Q), what qualifies as doing it better or worse, or ideally, and so on. The standards may be

explicitly statable or partly statable. They may also be exhibited in the consistency of judgments in regard to "better" or "worse" etc.

3.2 There is a significant conceptual and methodological gap between evaluation based on a task analysis and the rigorous evaluation, described in 2.0, of the contribution of an individual to the functioning of a social system.

3.2.1 In principle, an indefinitely large number of different task analyses of the functioning of a social system are possible. Different task analyses will in general differ substantially or even radically in regard to their adequacy (e.g., their coherence, articulation, completeness, etc.). Thus, the adequacy of an evaluation along the lines of "How well is the individual doing the assigned job?" is strongly limited by the adequacy of the task analysis which generated the job description. Most task analyses specify only the most readily definable or observable requirements, leaving the rest implicit or subject to "the requirements of the situation."

3.2.2 It is quite possible for every individual within a social system to be "doing his job" adequately as defined by the standards for that job and still have a poorly functioning social system. This is because the analysis of SS functioning into discrete jobs X_i , Y_i , and Q_i in general sacrifices the articulation of the possible ways of doing each of these jobs, so that various sorts of mismatch between P's behavior and that of R and Z are possible even when each is "doing his job" (in one of the ways it can be done).

For example, if the job of providing type K information to Z is P's job, P may produce the information in the form of a string of

I's and O's and be "doing the job" adequately, in a formal sense (i.e., the information is there). In such cases, we tend to introduce additional 'requirements' such as "in convenient form," "in a timely manner," "without a lot of useless information thrown in," and so on. Such 'requirements' may be seen as double negatives (Appendix D) which amount to "doing the job of providing type K information to Z without going wrong in this way, namely, providing it in incomprehensible form or in this way, namely, without taking too long about it, etc."

Note, however, that "not unintelligible," "in a timely manner," and so on lead to the corresponding questions "Well, what is intelligible enough?" "How quick is 'timely' enough?" and so on. These questions in turn lead back to the basic framework of directly evaluating the difference it makes in SS functioning, though there are ways of approximating these results on the basis of human judgments.

Just as there is an indefinitely large number of ways of doing a task analysis for a given SS, there is an indefinitely large number of ways in which doing job Q can go wrong, and we cannot take infinite precautions against them (Appendix B, Precaution Paradigm). Actually identifying possible ways of going wrong and safeguarding against them is an approximate procedure, and it will not guarantee that things go right.

- 3.2.3 The problem of sacrificing functional articulation by analyzing a process into a number of discrete "jobs" is not peculiar to information systems or to social systems. It is inherent in the nature of a

task analysis as a limiting case of process description (Ossorio, 1978, p. 59). For an example at the level of individual behavior, consider the following dialogue.

Wil: Say, Gil, you know how to play Bridge, don't you?

Gil: No.

Wil: Now wait a while. Surely you know what an ordinary deck of playing cards is.

Gil: Yes.

Wil: And you can distinguish among the four suits--clubs, diamonds, hearts, and spades?

Gil: Yes.

Wil: And you can distinguish different values, such as Ace, King, Queen, et cetera?

Gil: Yes.

Wil: And you can shuffle a deck of cards?

Gil: Yes.

Wil: And you can deal cards to players?

Gil: Yes.

Wil: And you can take a trick with a trump or a higher card?

Gil: Yes.

...

...

...

Wil: Well, now, look--there's nothing that gets done in a game of Bridge that you haven't said you know how to do. So what do you mean, you don't know how to play Bridge?

Since the sequencing, the timing, the contingencies, and so on are crucial for playing Bridge, it is obvious that Gil may be a disaster at the Bridge table even though he can do adequately every single "job" involved in the playing of Bridge.

3.2.4 The problem of articulating function across jobs is a general one. A special case of the problem of articulating/coordinating the possible ways of Z doing job Q and P doing job X arises when P is an information system and Z is a human user, or vice versa. These problems are conventionally designated as "human-computer interface problems." Other cases of interest arise when both P and Z are information systems or when P is an analyst and Z is a decision maker. The in principle solution is to embed the job descriptions in the corresponding process descriptions (task analyses are incomplete process descriptions). This is why the conceptual model in Section 2.0 is couched in terms of social system functioning and why it is methodologically critical to have a systematic way of representing processes and object-process configurations.

3.3 Nevertheless, it is possible to evaluate on the basis of job descriptions. Indeed, this is the primary way in which we identify an individual who is doing well in a poorly functioning organization. The "requirements" toward which an information system is designed and developed are, in effect, the job description for the IS within the social system. In light of the preceding, we can distinguish between "success" and "appropriateness" as criteria for evaluation.

3.3.1 The question of whether a given individual is doing a given job successfully is decided by reference to the functioning of the social

system in which the individual participates and of which the job is a structural part. If the social system is functioning normatively or within acceptable limits, then all of its components are functioning successfully. Relevant data for this evaluation are data concerning social system functioning.

The effectiveness of social system data is increased, potentially to a decisive degree, if it serves hypothesis testing concerning system functioning rather than merely establishes descriptive facts. (See Appendix A, Appendix B, and the "simulation" portion of the procedural model in Section 4.0.)

- 3.3.2 The question of whether a given individual is doing a given job appropriately is decided by reference to the job description or performance requirements. Where the job in question is that of an information system the relevant data for this evaluation are primarily data concerning the interaction of the IS with a set of human Users. This is because an information system as such has no executive functions and therefore takes no independent actions the results of which could be examined independently of Users, so that its significant function is to mesh with the activities of the Users. Here, the relevant data will consist explicitly of human judgments (e.g., on ease of use, comprehensibility, timeliness, etc.) or will depend indirectly on human judgments (e.g., the "objective" ratio of relevant documents to number of documents retrieved).
- 3.3.3 Ideally, "appropriate" functioning, in the sense of satisfying job requirements, is both explanatory and, in a sense, predictive with respect to "successful" functioning. This follows from the fact

that, ideally, jobs and job descriptions represent an adequate task analysis of social system functioning. If successful social system functioning (in its real world context) can be analyzed into the doing of tasks Q, X, and Y, then the doing of these tasks without going wrong in the ways specified by the (double negative) additional job requirements would both explain successful social system functioning (if in fact it was successful) and be empirically predictive of it.

The standards of appropriateness for a given job may be taken to reflect more or less implicit, unsystematic, or informal analytic and explanatory efforts of this kind on the part of interested or knowledgeable individuals or groups. Where judgments of this kind are available and accepted as valid, field testing for appropriateness of job performance can serve as a limited substitute for the criterion of social system contribution.

- 3.3.4 A similar relation holds between having certain individual characteristics and doing the job appropriately or successfully. If the job description is "provide type K information in a timely fashion," a task analysis may lead to the reformulation of this task as "(a) Receive type M basic facts and (b) output either these or any truth functional derivative facts (c) upon request in Query Language L (d) with an average response time of eight seconds. (Any description of the hardware or internal functioning of the IS will be an individual characteristic description.)

If the task analysis is apt, then providing truth functions of type M facts with an average response time of eight seconds will

be explanatory and empirically predictive of appropriate job performance.

- 3.3.5 In this connection it may be noted that, historically, one of the major limitations of real world task analyses has been that they tend to be seriously incomplete. Because they do not account for all the relevant aspects of the task either in the positive subtask structure or in the double negative hedges, they tend to fall seriously short in regard to both prediction and explanation.

Since task analyses are essentially elliptic forms of means-end descriptions and means-end descriptions are elliptic forms of process description, it is not surprising that task analyses tend to be incomplete for predictive and explanatory purposes.

- 3.3.6 Still, knowledgeable observers can and do make judgments to the effect that an IS with certain characteristics will perform a given job with roughly a certain degree of success or a certain likelihood of success, either overall or in certain respects. (For example, a variation on this procedure, referenced to an implicit range of tasks rather than explicitly to given tasks, was used in selecting the configuration for the Military Family of Computers (1977).)

It would be possible to justify the evaluation of an IS on the basis of its individual characteristics under either of the following conditions.

- (a) One might be able to show that human judges could accurately transform individual characteristics descriptions (e.g., "average response time of eight seconds for truth functions of type M facts") into evaluative ratings of IS job performance

(appropriateness). The criterion would be independent ratings for appropriateness of job performance based on observation.

One would have to show some degree of empirical generalization for the results, however, since the process of validating the use of individual characteristics involves collecting appropriateness data and for the system in question the appropriateness data would be preferable to the individual characteristic data.

(b) One might be able to show that there was a consistent mapping function or set of functions connecting individual characteristic descriptions and job performance evaluations (either global or componential).

In the simplest sort of case, for example, one might show that "timeliness" is a highly nonlinear but essentially monotonic function of response time. It is unlikely, however, that such mapping functions would in general be mathematical ones.

Like the preceding program, this program would require observational ratings of appropriateness of job performance. The gain of the function/mapping approach over the preceding ad hoc demonstration is essentially the same as the gain of quantitative evaluation over merely comparative evaluation. The function/mapping approach is a major ingredient in the Simulation Paradigm.

3.3.7 Presumably, the current design of information systems represents the more or less informal application of the kind of knowledge and empirical associations described above. Correspondingly, where either the relevant knowledgeability or empirical associations are

available, it becomes possible to perform an evaluation of an information system by bench testing to establish its relevant individual characteristics as a limited substitute for field testing to establish its satisfaction of job requirements.

3.4 We may distinguish three "levels of description," corresponding to three sizes of context, which can be used for the evaluation of information systems.

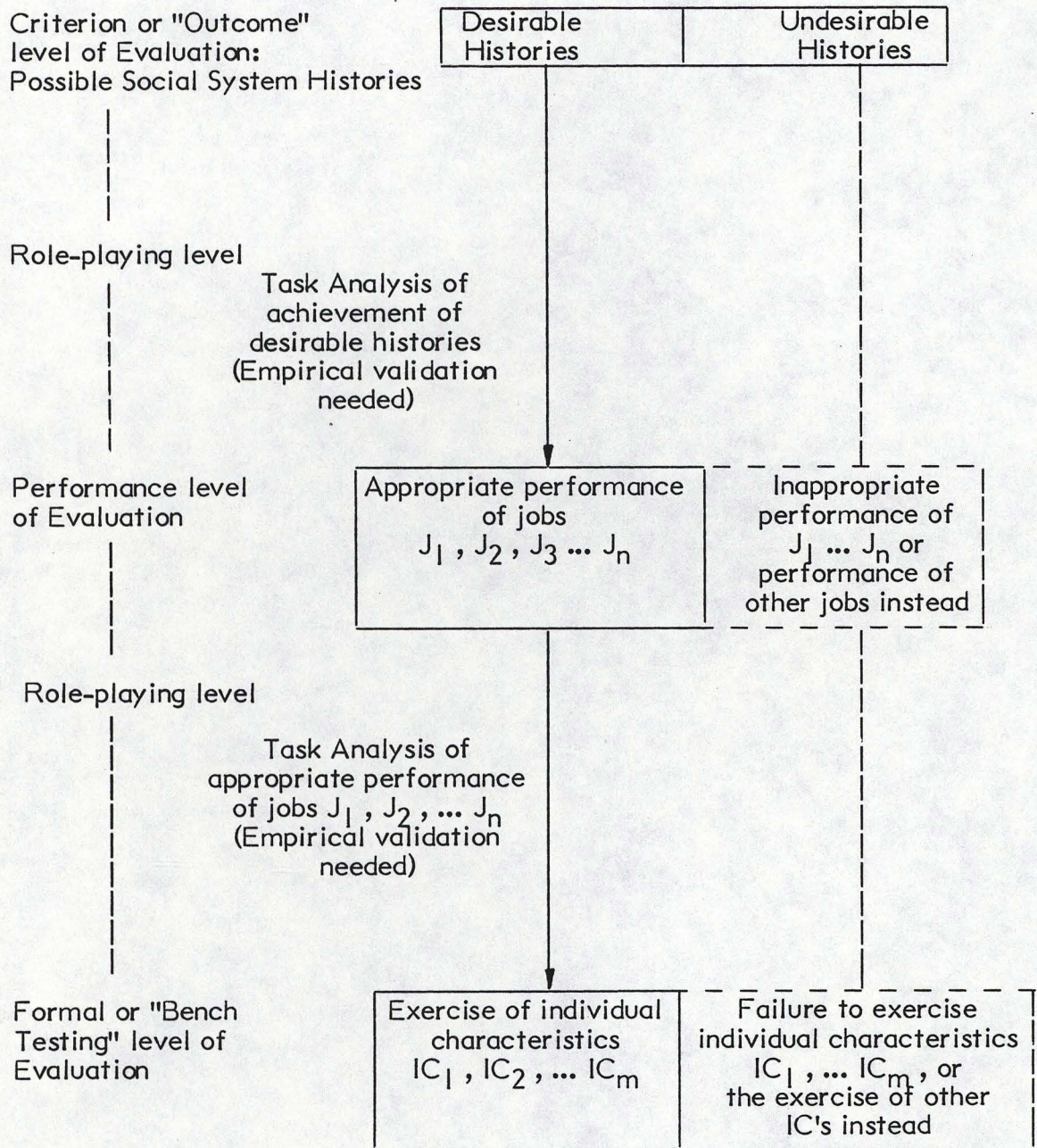
- (a) The "formal level" refers to the exercise of the characteristics of the system in question. This is the "bench testing" level.
- (b) The "performance level" refers to the doing of the job (more or less appropriately) and includes as a special case interacting appropriately with a User.
- (c) The "outcome" or "criterion" level refers to the contribution of the IS to the functioning of the social system and includes as a special case the successfulness of job performance.

The levels are hierarchically related, with the performance level reflecting a more or less implicit task analysis of the outcome level and the formal level reflecting a more or less implicit task analysis of the performance level. These relationships are shown schematically in Figure 5.

From the standpoint of evaluation the levels in the hierarchy are directly related to validity and inversely related to convenience, accessibility, and practicability, so that we have a tradeoff situation.

The general limitations on the adequacy of task analyses place strong limits on the advisability of depending on the more convenient methods as a substitute for the more valid ones. Conversely, direct

Figure 5. Levels of Description and Evaluation



observation of the criterion, i.e., the possible histories of the social system, is simply not possible. These considerations provide boundary conditions for the formulation of a technical/procedural approach to the evaluation of information systems.

4.0 A Procedural Model for Evaluation

In this section we will make reference to three experimental paradigms, i.e., the ANOVA (analysis of variance) paradigm, the Precaution Paradigm, and the Simulation Paradigm. More extensive discussions of these are found in Appendix A and Appendix B. The following are brief synopses.

- (a) In the ANOVA paradigm we do a parametric analysis of the phenomenon and systematically vary the values of the parameters, singly and in combination in order to determine whether differences in these values are associated with differences in some dependent variable which is of interest, e.g., an evaluative index for an information system.
- (b) In the Precaution Paradigm we conceptualize an enterprise in a real world context (e.g., operating a given information system in a given social system). We then conceptualize and prioritize the major ways that the enterprise could fail, be deficient, or be successful. We then formulate which events or states of affairs would be observable and accessible to us if the enterprise were to fail or be deficient or be successful. The deliberate making of these observations provides either a warning that it is going wrong in the hypothesized way or a reassurance that it is probably not going wrong in the way we feared or that it probably is going right in the way we were concerned about. The Precaution Paradigm allows us to take these precautions in order of priority, i.e., in the order of our degree of concern and also to limit our precaution taking to those possibilities where we have a genuine

practical concern (as against merely recognizing abstract logical possibilities of going wrong).

- (c) In the Simulation Paradigm we model the phenomenon of interest using existing resources for real world representation. In the case of evaluating information systems, this would involve Process Description which gives us a structure-function representation of the functioning of the IS and the social system of which the IS is a part. The Process Representation involves a Contingency analysis which specifies which course the process takes at a given point as a function of the relevant events or states of affairs. It is the explicit formulation of this functional analysis which permits the simulation to "run." In the typical use of the Simulation Paradigm we assess the contingencies and predict the observable course of events. Since in principle the entire set of functional relations is tested by every prediction the use of the Simulation Paradigm allows us to build up a strong, empirically based confidence in the stated functional relationships on the basis of a surprisingly small amount of data. (In this respect it is like classic theory-testing research, but then, the model including the functional analysis is, in effect, a theory of the phenomenon.) One of the primary virtues of the Simulation Paradigm is that it permits us to estimate or approximate phenomena to which we have no direct observational access (e.g., possible histories of the social system).

Each of the experimental paradigms is a conceptual framework which provides a distinctive rationale for collectiong data, using it, and acting on the results.

The Precaution Paradigm provides the most coherent overall framework for the problem at hand. Our aim is to evaluate an information system and to take precautions against going wrong either (a) by investing too much in a full-scale validation of a system which could easily have been shown to have decisive limitations or (b) by developing excessive confidence in a system on the basis of merely convenient data gathering when misplaced confidence could well have unfortunate or even disastrous consequences.

- 4.1 The procedural model involves a paradigmatic sequence of evaluation procedures. The sequence begins with procedures involving maximum convenience, minimum cost, and maximum accessibility of the phenomenon of interest. The subsequent movement is increasingly to procedures involving greater cost and/or less convenience and/or less accessibility (the three tend to be very strongly tied), but also greater validity in principle.
- 4.2 Following this sequence from the starting point is essentially precaution (a), above. It provides a practical guarantee that the more easily recognized flaws and deficiencies will be identified earlier and that validation is not a wasted effort. Conversely, being able to follow the sequence to its end point is essentially precaution (b), above. It appears to provide the greatest assurance currently available that prior validation is available in the areas of greatest concern.
 - 4.2.1 The first level of testing is bench testing to assess individual characteristics of the IS which are assumed to be relevant. At this level, the ANOVA paradigm appears to be the most natural one to use, possibly in connection with a simulation capability.

Among the individual characteristics which are likely to be relevant are (a) component reliability, (b) peak load capacity, (c) response time for representative tasks, and (d) memory capacity.

- 4.2.2 The second level of testing is at the level of job performance with experimental Users, a hypothetical social system, and a more or less realistic job description.

This setting provides some opportunities for bench testing methods, since the hypothetical aspects can be changed at will. However, the natural framework for much of the 'human factors' aspects is the precaution paradigm.

Among the procedural characteristics which are likely to be of interest are (a) ease of access, (b) intelligibility of output, (c) relevance of data retrieved, (d) precision/accuracy for retrieval, (e) timeliness of response, and (f) overall ease of use.

- 4.2.3 The third level of testing is field testing with actual Users in an operational setting. From this level on, manipulatory power is minimal but observation across naturally occurring variations in circumstances is available. ANOVA techniques become unworkable, and the indicated methodology is the Precaution Paradigm.

The archetypal logic of the Precaution Paradigm is an error-detecting feedback loop: "Let's do it, but let's see if it goes wrong in a way that we are concerned about." An elaboration is "Let's do it this way rather than that way to make it less likely that it goes wrong in this way."

As an initial precaution against overconfidence or hidden defects, the IS should be introduced as an additional alternative if it is designed to replace another IS or as an experimental system if it is not. The precaution provides a functional insulation so that the IS can be normally operative without the social system suffering the full impact of its defects if there are any serious ones.

Identifying serious (enough) ways of going wrong is the responsibility of the evaluator. In general, they would be identified by consideration of the social system and job requirements and by consultation with Users and other interested parties. The point is that if the operation of the IS goes wrong in its real world functional setting, then it is possible to make observations which directly establish that it has (or has not) gone wrong in a particular way or which provide evidence for one or the other conclusion. Field testing with actual Users has a validity which experimental testing with role-playing Users (level 2) does not.

Testing at this level would be most likely to involve either the primary validity of the information or various double-negative restrictions or concerns. For example:

1. It doesn't give me the information I really need. (E.g., "It tells me something about what they are doing, but I really need to know what progress they are making, and that I have to guess at.")
2. It doesn't distinguish between hard information and subjective estimates, and that makes a lot of difference to me.

3. It's a bother to use because it takes so long to respond that even when I get good data I lose my train of thought.
4. It doesn't give me all it has that I want.
5. It gives me so much junk to sort through to find what I want that I don't bother unless it's really important.
6. It's a real pain to have to play "20 questions" when all I want is a simple answer to a simple question.
7. Reading it off the display I make more errors than I can really afford.
8. A lot of garbage got into the data base last month and no one knows how much difference that's still making.
9. When I've gone without sleep the night before I have a tendency to ask simpler questions than I should and not follow up on them the way I should.
10. If I had a little systematic training I would have a lot easier time using that system.

The methodology of the Precaution Paradigm is primarily an observational methodology dealing with actual occurrences. As noted above, it extends to imaginary occurrences with foreseeable outcomes, risks, or other characteristics. A logical extension of the latter is the concern about unacceptable outcomes which are not readily foreseeable in part because the circumstances under which they might well occur are currently hypothetical and therefore not observable. In the present context, the most obvious example is "But will it hold up under battle conditions (or this or

that disaster scenario)?" It is this concern which leads to the final level of evaluation.

- 4.2.4 The fourth level of evaluation is modeling and predictive hypothesis testing. The purpose is to predict the functioning of the IS and the social system under conditions which are hypothetical but also of interest. The appropriate methodology for this purpose is provided by the Simulation Paradigm. The hypothetical histories which are modeled can then be acted upon in accordance with the Precaution Paradigm.

It is in connection with the modeling effort that a systematic and substantively adequate way of representing the social system, as contrasted with a merely nominal or abstract symbolic representation, is essential. Successful modeling in the fashion envisioned here would involve the following. ("Wartime functioning" is used as a paradigm example of the hypothetical conditions which it is of interest to model.)

- (a) The identification of the parameters of the social system which would be most relevant to the difference between "normal" and "wartime" functioning. These are designated as Contributory Parameters.
- (b) The definition of values taken by each parameter
- (c) (Optional) A task analysis defining the expected/required contribution of the IS to the social system.
- (d) Methods for assessing the contribution of the IS to the SS functioning under "normal" field conditions (these are the methods discussed above and assumed here).

- (e) Evaluative data for the IS under some range of values of the Contributory Parameters under existing (presumably, more or less "normal") field conditions.
- (f) The construction of an explicit simulation model of the SS including the IS. The simulation model would include a representation of the objects, relationships, processes, events, and states of affairs which would constitute the possible histories of the SS. It would also include a representation of the contingency structure of the system. This would be given by specifying explicit functional dependencies of historical possibilities on particular events or states of affairs, which are designated as Contingencies. Among these Contingencies would be some parametric values of the Contributory Parameters.
- (g) The initial validation (and modification) of the simulation model based on the existing data ((e), above). Subsequent validation by predicting new data based on the contingency structure.
- (h) The specification of relevant scenarios for the hypothetical "wartime conditions" of the social system.
- (i) The translation (if needed) of the scenarios into a specification of the parametric values of the Contributory Parameters.
- (j) "Running the simulation model under these parametric values.
- (k) The definition of Version Classes of social system functioning by reference to relevant scenarios.

- (l) The mapping of the simulation results onto the relevant Version Classes for evaluation purposes. In the simplest case, for example, all of the scenario simulations will fall into "acceptable" or "successful" Version Classes, in which case there is no grounds for worry (or, conversely, they will all fall into "failure" or "unacceptable" Version Classes, in which case some corrective action is called for). In general, some distribution is to be expected (1) from a probability standpoint into "successful" and "unsuccessful" Version Classes and (2) from a qualitative standpoint into the various relevant Version Classes.
- (m) Using the results described above, it would be possible to do further systematic simulation studies varying relevant parametric values in order to identify the borderline, or "limits of tolerance," parametric values. These would serve practically to define the limits of the system's capability.

5.0 An Examination of a Contemporary Information System

In this section we examine METER, a recently developed information system, in light of the conceptual and procedural models.

5.1 A Summary Description of the METER System

The Message Extraction Through Estimated Relevance (METER) Project was developed under RADC contract F30602-77-C-0123 by the Pattern Analysis and Recognition Corporation. It is a fully automatic information storage and retrieval system supporting User access by natural language query to a large collection of documents in the form of unformatted English text. Although in the initial experimental project METER was significantly limited by time and space as well as the storage and processing capacity of the operational hardware (DEC PDP-11/45) selected for the project, it was designed to keep pace with the flow of incoming messages, to store up to 30,000 of the most recent messages, and to be usable by military personnel having little experience with computer information systems.

METER was designed for a high degree of integration between the computer and User, as indicated by the following overview comment: "The successful exploitation of large amounts of textual data in an online information system requires a high degree of cooperation between an end User and a computer. The job of a User in an information system is to provide the insights and to ask the right questions of the system. The job of the computer is to perform all of the routine analysis and comparison of stored data. In an effective information system the interaction between the User and the computer must make

it easy for the User both to ask the necessary questions and to interpret the computer's response."

The implementation of this statement in METER was accomplished by a number of strategic subroutines, many of them necessitated by the selection of an analytic paradigm which is in many ways alien to the typical thought process adopted by a naive User to perform the same analytic task. The procedure in question was based on the use of a statistical retrieval methodology. This methodology has two major components, i.e., a text processor to transform messages as they arrive into vector form and collect statistics on them and a content analyzer which updates the retrieval algorithms on the basis of collected statistics. A retrieval component coupled with message statistics rounds out the system.

Incoming messages are normalized, the suffixes are removed, and the text is reduced by the text processor. A microconcordance is developed for each message and statistical information on the frequency of each term is collected for the message. Overall, the message term statistics are collected, content terms are collected, and term to term associations are collected by the content analyzer. The files produced by these procedures serve as the basic access algorithms. Access (retrieval) is built on a system of previous message analysis. It accepts a natural language query from the User and returns a list of terms ordered by relevance as determined by a statistical procedure. The User of the system can, and for the most part does, remain unaware of the inner workings of the system and deals only with the User interface.

The interface was developed on the basis of feedback from individuals considered to be representative of target end Users. A minimum criterion for the interface was that it should provide an improvement over the common Boolean key word systems which require that the end User be trained in this form of logic and be familiar with the associative format and content of the data base.

An example of changes based on User feedback was the shift from a set of numbers which were difficult to interpret to histograms which carried the relevant information and were easy to interpret. Using the histograms, the Users were able to make judgments that resulted in more satisfactory retrieval.

User feedback also resulted in a change designed to shorten the time required to display an entire retrieved message on output. This change involved a major shift in the ways that text and queries were processed. The shift involved limiting the number of key index terms needed for document retrieval and subsequently displayed. This display also served the purpose of making the retrieval process less mysterious and allowed the User to use the system in a more informed fashion.

Other changes based on feedback from a pilot demonstration were a change in the associative schema selected to ensure system flexibility and a change in the simplicity of the scheme whereby a naive User could access the system in a competent and useful fashion; the resulting system of six commands allows users to identify themselves, enter queries, and retrieve and display text on their terminals.

Queries are normally entered in natural language with the critical parameter being the spelling of each word. No format is anticipated so

that punctuation, sentence structure, and so forth are not used by the system. METER allows the User to specify the number of documents to be retrieved and these are retrieved in accordance with their computed order of relevance to the query. Retrieval relevance was found to increase as the amount of detail in the query increased.

One interactive feature of METER allows the User to examine the retrieved documents and make judgments as to their relevance. The system then uses these judgments in addition to the original query to perform a new retrieval. It also permits access to the system using a limited Boolean logic for those Users who are familiar with the system content or are looking for specific content.

Offline aids enable Users to learn about the text collection environment, the data base content, and the statistical procedures used to store and retrieve text. These aids enable the User to approach the system in a more knowledgeable and effective way but are not requirements for using the system.

Another feature of the system is its ability to store, retrieve, and utilize query data associated with individual Users. This enables METER to orient and respond to individual Users based on their history of interaction with the system.

5.2 Review of METER

By way of review we may note the following.

I. METER was designed to meet the following requirements.

- (a) To provide access to information contained in large, dynamic, text data bases, specifically,
 - (1) a five day message file

- (2) updated daily
- (3) with about 6,000 new messages daily
- (b) for purposes of warning, planning, and decision making
- (c) using current standard minicomputer hardware and software.

Additional requirements were

- (d) Response time to queries is less than one minute for five simultaneous users
- (e) Users do not require extensive knowledge of computers or their use
- (f) Documents are available for retrieval within 15 minutes of arrival.

Among these, (a) appears to be the basic job requirement, with (c), (d), (e), and (f) being double- negative precautionary restrictions and (b) being interpretable as another double negative or as an elaboration or refinement of (a).

2. Requirement (a) defines a (more or less heterogeneous) class of jobs rather than an actual job. We may imagine a variety of social systems employing a variety of information systems all importantly different from one another and all fitting the description (a). Correspondingly, we may regard the METER prototype either as an actual system awaiting completion or as a methodology and an exemplary embodiment for accomplishing (a). Although there is some interest in each we will adopt the first unless specifically noted.

3. The METER prototype involves a task analysis of (a), namely (approximately) "by using statistical association methods on significant terms to match natural language queries with documents." The further task analysis of this task into its detailed implementation was tentative and subject to revision. The approach taken was an extension of earlier methods for retrospective document searches. A variety of alternative task analyses were possible,

including classical key-word methods or a Classification Space approach (Ossorio, 1964) which METER significantly resembles.

4. Of the initial double-negative constraints, (c), (d), and (f) are clearly at the level of individual characteristics and (e) is formally of this sort although there is a clear implication concerning interactions with Uses. In contrast, the developmental work on the prototype, including the extensive work on the User-friendly interface and User support, consists almost entirely of precautions against going wrong in various ways at the level of appropriate job performance.

In effect the process of development was also the process of evaluating, in specific ways, at Level 1 (bench testing, individual characteristics) and at Level 2 (appropriate job performance, role-playing Users). Evaluation at Level 3 (actual Users) is indicated as the next step. Some of the evaluation procedures described below would have been possible at Level 2.

5.3 Prospective Evaluation of METER

The description of the METER prototype presented by Brown, et al. (1979) could be expected to raise some number of questions for the critical evaluator. Among them are the following.

- (a) Does the system have the information wanted by the User?
- (b) If it does, will the User get the information?
- (c) Does the system have drawbacks which significantly nullify its advantages?
- (d) Is the system better than alternative systems for the purpose at hand?

- (e) Does the information the User gets make an important difference?
- (f) Would the User be any worse off without the system entirely?

Such questions, their grounds, and possible ways of answering them are discussed below.

5.3.1 In spite of the reported previous evaluative work basic validity questions would arise for several reasons.

- (a) The authors are on shaky ground in arguing that results count and theory doesn't. This argument is fully compelling only at the level of direct evaluation of social system contribution, i.e., Level 3 and Level 4 evaluation. Level 1 and Level 2 evaluation, which is all that is reported, require more than merely empirical results in order to provide some expectation that the real world results will be comparable to the experimental results.
- (b) On the whole, sampling theory does not provide a dependable basis for generalizing in situations such as the present one. The choice of Users and experimental data bases as similar to the projected Users and data bases qualifies only as an elementary precaution. This is different from controlled random sampling from a population.
- (c) Given the extensive empirical tailoring to the experimental Users, there is some ground for expecting that extensive tailoring would also be required for a set of real Users. General experience with empirical methods indicates that extensive ad hoc tailoring to a given situation or group

generally produces results which are spuriously good in that they do not generalize well to comparable situations or groups.

- (d) Since the distinctiveness of terminology dealing with certain subject matters (e.g., scientific, technical, and some politics and games) is greater than for other subject matters, the ease of discriminating natural language messages would be expected to vary with the content of the data base. Since the nature of the real world data base and its relation to the experimental data base is left open there is some question about the generalizability of the Level 2 results. There is some unknown level of expectation that the "dynamic data base" will in general have the desired information at all and that when it does not the User will be able to find that out readily.

Given the availability of User judgment, it is a simple matter to establish the proportion of queries for which the information wanted is present in the system: (a) Successful retrieval provides positive cases and the history of retrieval success will set a lower bound on the looked for proportion. (b) The ideal estimate would require examination of the entire data base in cases of unsuccessful retrieval. This procedure could be made considerably more economical by processing the data base for a whole group of queries in one pass. One could also sample a portion of the data base and estimate accordingly.

5.3.2 One of the principal questions which arises for document retrieval, as distinct from factual responses to factual queries, is the question of the false positive and false negative rates. The traditional indices in this respect are precision and accuracy. Moreover, there is no compelling substitute for this kind of validation data. The report by Brown et al. mentions some limited data of this kind, but no figures are given. In any case, the questions raised about the generalizability of the Level 2 evaluation call for new data from real Users and an actual information system within a specific social system.

Given the availability of User judgment, ideally, precision and accuracy are calculated on the basis of retrievals and examination of the entire data base. Where this is impractical because of the size of the data base, as would be the case in the anticipated use of METER, it is quite feasible to work only with a known percentage of the data base obtained through random or systematic sampling.

5.3.3 The question concerning drawbacks is an umbrella question which requires the specification and evaluation of particular drawbacks. For example, the precision and accuracy measures already get at the cost to the User of getting what he does get. Other "drawback" aspects would include

- (a) It is too much trouble to read through all those entire messages (even when I do get the right one) to get the information I want.
- (b) It is too confusing to operate the query facility.

- (c) I never know how to formulate what I want to know so that I get a good answer.
- (d) It takes too much time--it's frustrating.
- (e) I get contradictory answers from different messages.

All of these are questions of appropriateness of job performance and the relevant data are direct evaluative judgments on the part of the Users.

5.3.4 Particularly if the preceding procedures result in significantly negative evaluations, there would be some interest in a comparative evaluation of the METER prototype as one alternative among several.

- (a) In general the comparative evaluation is accomplished by performing the same evaluations (e.g., the preceding) independently on both alternatives and comparing the results.
- (b) Two plausible alternatives are (1) a Classifical Space retrieval (Ossorio, 1965, 1966) and (2) a combination of associative retrieval (METER) and judgment based retrieval (C-Space). The C-Space approach is plausible because it resembles METER in almost all aspects except the indexing principle (it represents an alternative task analysis). Specifically, it accepts natural language queries and applies the same indexing to queries as to documents; it indexes documents within a vector space on the basis of the distinctive words which appear in the document and retrieves documents in order or relevance; it allows for User feedback to refine the retrieval process, is completely automatic, and allows

various types of User support. Complementarity between the two lies in that (1) The C-Space approach has a strong theoretical basis and almost certainly has significantly better precision and accuracy figures and a stronger empirical track record in Level 2 assessment, whereas (2) METER is almost certainly more completely open-ended with respect to changing content in a dynamic data base.

- 5.3.5 The question of whether the information provided by the system makes an important difference (in social system functioning) is raised in part because METER is apparently projected for use by analysts who will be casual users and have gotten along without it heretofore. The question may be interpreted as a Level 3 problem or as a Level 4 problem. We will approach it as a Level 3 problem.

Consider the following scenario: In the course of a routine retrieval User A is denied the documents indicated by the system to be most relevant. He indicates that he did not get the desired information, and continues about his business. At a later time designated by a supervisor or other monitor he is given the previously denied documents and indicates that he did get the desired information. He now goes back to the previous task and handles it in the light of the new information. What he does differently with and without the information is observable and also open to self report. The expected consequences of this difference are described by User A (this is the Level 4 aspect of the problem) or by another, comparable User, User B. The value of these consequences is evaluated by User B. The comparison of these values

provides the index of the contribution made by having the information in question. Given a random or systematic sampling of retrieval requests to be handled in this way, a relatively small number of replications of this procedure could be expected to produce practically useful results for a general conclusion about the utility of the system.

Or an alternative scenario: User A makes a query and indicates that he got the information he wanted (one could combine data for all the information relevant to his project). After he has made use of the information he is queried as to what he would have done had he not obtained that information from METER. The expected consequences of both the actual and hypothetical actions by User A are described by User A or User B. The remainder of the scenario is identical to the preceding one.

A variation would be to ask "How would you have handled your project before METER was introduced?"

Of course, there are other possibilities.

- 5.3.6 The precaution aspect of the preceding scenarios is the precaution against going to a lot of effort and expense without any ultimate payoff (the overconfidence problem). The procedures described require the active cooperation of Users in a role-playing capacity. (Since they are real Users we regard this as a Level 3 evaluation.) Brown, et al. to the contrary notwithstanding, it does not appear that satisfactory evaluation of an information system such as METER can be accomplished without the active cooperation of the Users.

For example, the "unobtrusive" monitoring to see if the Users consistently use the system is relatively unconvincing as evidence of its usefulness. Convenience plus an occasional hit may be enough to generate consistent use. Going the last mile to use whatever resources are available in spite of infrequent hits may also generate a high level of consistent use. So also may the entertainment value of browsing. More likely, the Users may mistakenly believe they need the information.

As noted earlier, an initial precaution is to introduce a new information system either as an alternative or as an experimental innovation. In either case the User will have a legitimate interest in testing it out and if the social system is not seriously deficient at this point that should generate enough cooperation for evaluation purposes. (If the system is important enough, the evaluation is important enough, particularly if it is kept to a minimum by fully exploiting Level 1 and Level 2 possibilities.)

5.3.7 In light of the foregoing review and examination of METER based on the report by Brown et al., the following summary evaluative conclusions can be drawn.

A. Evaluation procedures which have been conducted are Level 1 and Level 2 procedures. These procedures have shown positive outcomes; the METER prototype meets the requirements listed above in 5.2.

B. Evaluation should be continued with Level 3 procedures and if possible, Level 4 procedures. The tests already completed do not warrant substantial confidence that implementation of METER in a

field setting would provide effective or efficient retrieval or would contribute significantly to an existing mission. Among the reasons for this conclusion are the following.

- (1) Lack of validation of the requirements
- (2) Prototype is excessively ad hoc for generalizing the test results to an operational setting
- (3) The marginal utility of METER over alternatives, including the alternative of no IS should be addressed
- (4) Methods do exist for informative and cost-effective Level 3 evaluation of METER and possibly for limited Level 4 evaluation also.

6.0 Future Research and Development

The survey of the information processing literature and the considerations emerging from the formulation of the evaluation models and the exercise with METER suggest that certain directions of future research and development would be more fruitful than most because their contribution if successful is relatively certain and there is good reason to believe that substantial success is likely. Some of these efforts have to do with the state of the art in the evaluation of information systems. Others have to do with the interface problem. Not surprisingly, there is some relation between the two.

6.1 Information System Evaluation

The state of the art in information system evaluation is such that particular evaluation efforts almost invariably look primitive, ad hoc, and subjective. In part this is because paradigmatic appraisal is completely context-bound by a historical context, is relative to a given person, and irreducibly requires human judgment. However, the same formulation which shows why this is so also shows in what sense empirical generalizations are possible, i.e., in descriptive appraisal. The procedural model shows where and why such empirical generalizations would be of considerable value for conducting information system evaluation. On the whole, the following appear to be among the most promising research and development directions to take.

- 6.1.1 Empirical generalizations are valuable because, within the limits of their empirical validity, they permit us to draw conclusions about the results of an evaluation without having to perform the

evaluation. (A similar value attaches to empirically validated conceptual models used in the Simulation Paradigm.) There are several types of descriptive appraisal which lend themselves to empirical generalizations.

- (a) In the area of appropriateness of job performance, evaluations of individual parameters can be mapped onto the overall evaluation of the system.

For example, for each different IS the factual basis for rating the IS as to categories or degrees of "ease of use" would in general be different. However, it might well be that across a wide range of systems, any one that was rated 2.3 or less on a standard 9-point scale was also rated unacceptable overall. In this case we would have a finding concerning Man the Computer User and also an empirical generalization which would allow the prediction of the overall evaluation under certain conditions.

For phenomena of this sort we would expect three ranges of values. The highest range would be associated with overall acceptability and the lowest would be associated with overall unacceptability. The middle range would be associated with various tradeoffs with other individual characteristics. For example, in this range low "ease of use" might be compensated for by high "accuracy of information."

- (b) Similarly, individual characteristics of information systems can be mapped onto (1) individual parameters of appropriateness of job performance, and (2) overall appropriateness of

job performance. Of these, the first seems most likely to produce viable empirical generalizations since one would expect less influence from other variables.

- (c) Likewise, all three of the foregoing (individual system characteristics, evaluation of individual appropriateness parameters, and evaluation of overall appropriateness of job performance) can be mapped onto degree of success in job performance (contribution to the social system).

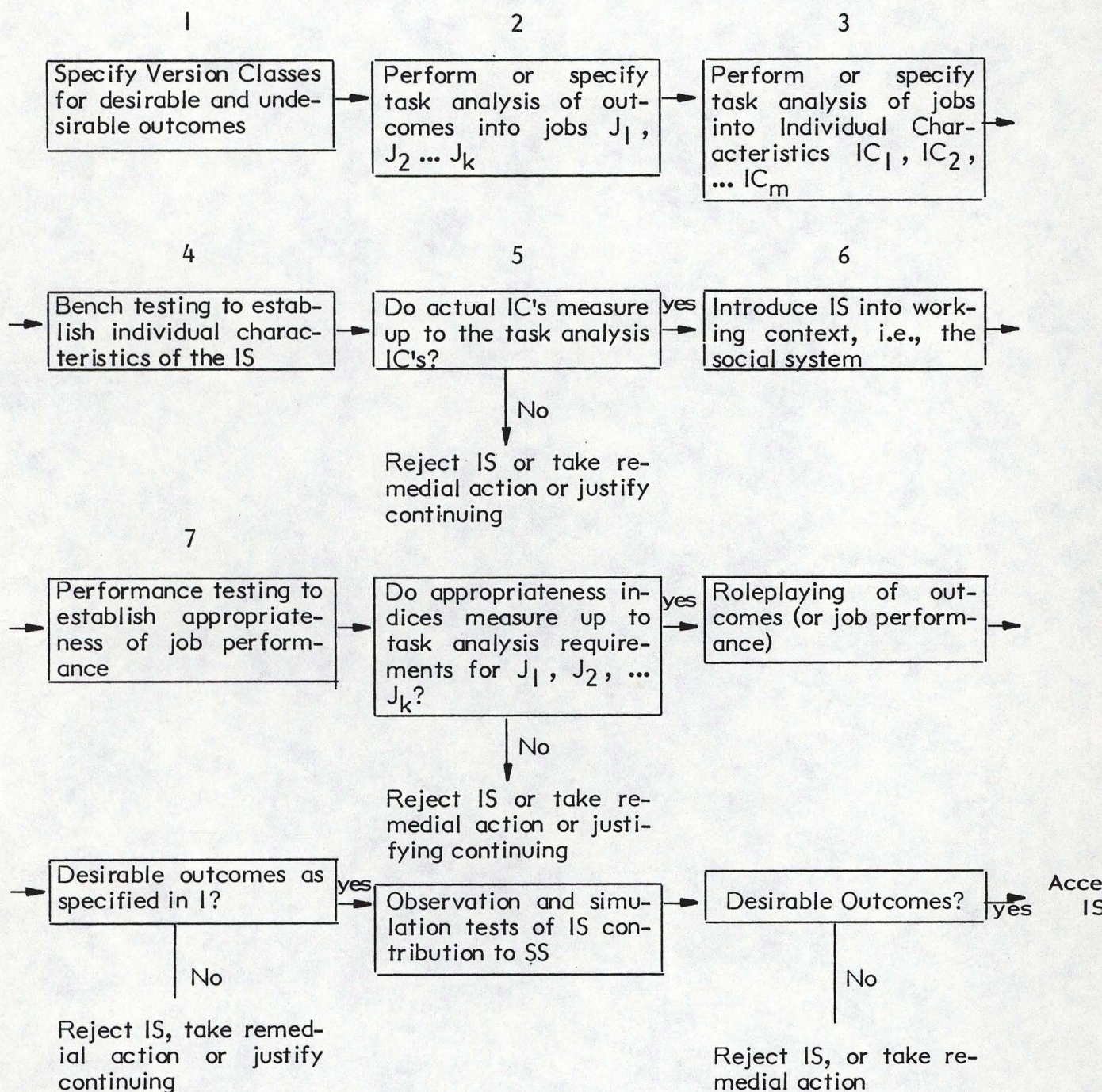
Figure 6 shows these relationships in diagrammatic form. Empirical generalizations of these kinds permit us to approach particular evaluations with significant resources in hand rather than starting completely new each time. They do, however, require a systematic program of research, since simply keeping a record of actual, ad hoc evaluations cannot be expected to provide the needed data.

6.1.2 Empirical research is needed to establish the moderator variables which have a bearing on the empirical generalizations described above and which serve to explain and predict exceptions or variations from the simple generalizations. Possible examples of significant variables of this sort are

- (a) Prior learning/training of the User
- (b) The nature of the task (social system function) performed
- (c) The mode of interaction between User and computer (question-and-answer, form-filling, etc.)

Although the results obtained from these studies are also empirical generalizations, they function as individuating principles

Figure 6. Flowchart of
Paradigmatic Sequence of Evaluation Procedures



rather than as aggregating principles. Both are needed in order to avoid the extremes of overgeneralization and complete idiosyncrasy.

6.1.3 Research is needed to develop in greater depth the essential background technology for Level 4 (simulation) evaluation. Relevant research would include the following.

- (a) Research to demonstrate or acquire the facility to represent adequately social systems of the kind involved in the use of information systems.
- (b) Research to demonstrate or programs to evolve standard Version Class definitions and implementations for the kinds of social system functioning under consideration.
- (c) Research to demonstrate successful predictive simulation of the kinds of social system under considerations, without regard to evaluation.

6.2 Interface Problems

We noted that work in the general area of human-computer interfacing fell into two general categories, i.e., (1) efforts to compensate for the difference between the data structure of the information system and the psychological structure of the User, and (2) efforts to reduce the difference by making the data structure conform more closely to the User or vice versa. The greatest part of this work falls into the first category. Given the seemingly endless body of techniques, programs, hardware, training, etc. produced by these efforts and their continuing lack of convergence or compelling success it does not appear that this is the most promising direction to pursue with an

eye to significant long term gains. The alternative has the advantage of parsimony and the potential for more powerful information processing.

6.2.1 If we interpret the term Artificial Intelligence as the effort to approximate human cognitive functioning by computer systems (as contrasted with the narrower effort to redescribe human cognitive functioning as the operation of computer programs) then data bases of the kind designed for artificial intelligence projects provide the markers for data base development designed to reduce the human-data base difference.

- (a) One of the key directions here is the development of representational power. It has been noted by a number of observers that general problem solving approaches have not produced impressive results in artificial intelligence. Rather, the most impressive results have come from sophisticated representations of a substantive domain (reflecting "expert knowledge"). A major research thrust in this genre, therefore, would be a sustained effort to increase our capability for real world representation. The efforts describe above in connection with simulation and Level 4 evaluation are directly of this kind.
- (b) It has long been a truism that the ideal form (or one ideal form) of human-computer interaction would be conversation in natural language. However, efforts in this direction have not been notably successful. There is some reason to believe that the problem here is a special case of the preceding one,

namely that significant progress will depend on more powerful substantial representation rather than content-free heuristics and merely syntactic or semantic analysis. In any case, interaction in natural language still appears to be a meaningful ideal to work towards.

- (c) Many of the current requirements on information systems include the requirement that they be usable by persons who do not have extensive training in the use of computers. This dimension seems clearly to be one with significant end effects. The degree of constraint on the rest of the system could be expected to go up sharply as the prior competence of the User approaches zero. This was one reason for mentioning it as a moderator variable above. It appears that a significant gain would be achieved on the whole if information systems routinely included a tutorial segment whereby a User could quickly upgrade his computer interaction skills to the (modest) point required to make use of the system.

6.3 Indications for I & W

The conceptual and procedural models presented above may be summarily redescribed as dealing with both positive and negative issues in information system evaluation.

The positive issues concern how one directly demonstrates positive value. In summary, this is accomplished by the following.

- (a) Social system representation which defines the criterial "successful outcome," or "desirable history."

(b) Hierarchical task analyses beginning with successful social system outcomes and ending with the internal operations of the information system or any intermediate point which is acceptable as "given." (See Figure 7.)

(c) Conceptual validation of the task analyses. For a given analysis this is accomplished by showing that the accomplishment of the analytically derived tasks could be expected to qualify as or result in the successful criterion outcome.

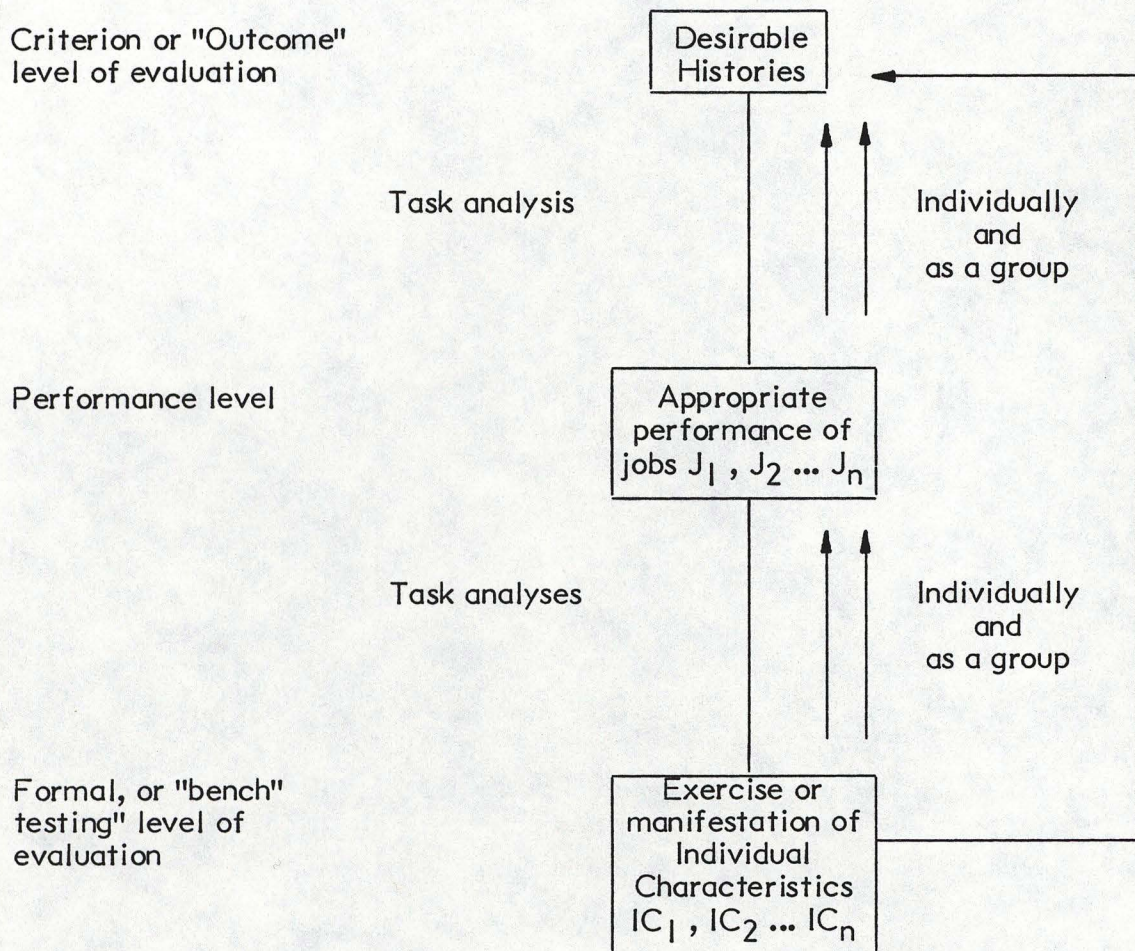
(d) Empirical validation of the task analyses. For a given analysis this is accomplished by showing that, at an acceptable level of probability or approximation, the accomplishment of the analytically derived tasks in fact does qualify as, or results in, the successful criterion outcome (recall the "Can you play Bridge?" example).

(e) Empirical validation of system components. This is accomplished by showing that the component (e.g., the information system) does perform appropriately the tasks defined by the task analysis.

The negative issues concern how one guards against being undermined or overwhelmed by flaws or limitations in the functioning of the system as a whole or of its components. The primary framework for doing this is provided by the Precaution Paradigm, in which various "disaster scenarios" are generated and dealt with in order of priority.

The resources presented above at the conceptual, methodological, and technical levels appear to have direct applications to the USAF I & W network. The most direct applications are described briefly below.

Figure 7. Empirical Generalizations



Arrows indicate possibilities for data gathering and empirical generalizations. In general, task analyses call for empirical validation. If the desirable outcome is analyzed as equivalent to or resulting from the appropriate performance of jobs $J_1, J_2 \dots J_n$, then it remains to be demonstrated empirically to what extent it is true that when these jobs are performed appropriately the desirable outcome is found.

6.3.1 Level 4 Efforts

It appears that the most critical long term application lies in the development of capabilities for Level 4 evaluation, i.e., evaluation at the criterion level of social system functioning. Present capability in this respect appears to be rudimentary in principle and unavailable in practice. The need is critical because Level 4 evaluation provides the substantive and methodological anchor for all IS evaluation in that it provides the criteria for other evaluation efforts dealing with either positive or negative issues as defined above. Other evaluation efforts are ultimately valid only insofar as they approximate the results of a Level 4 evaluation.

The development of Level 4 capability would include the following efforts.

A. Identify relevant social systems

As noted above, the "social system" referred to in connection with social system functioning is a meaningful functional unit such as a task force, a military base, a systems command, an experimental project, and so on. Such units will be hierarchically ordered and in general the "social system" to which the information system contributes may be any one of a number of distinguishable functional units. Although any of these could be used for evaluation purposes, the obvious place to begin is with social systems which are mission oriented and closest (in the hierarchy) to the information systems to be evaluated.

B. Represent social systems

The systematic representation of social systems in a form adequate to the evaluation task is a key capability which needs development. The Object/Process (structure/function) and Task Analysis formats described in "What Actually Happens" (Ossorio, 1978) and summarized in Appendix B of the present report have been used to construct effective representations of business and law enforcement organizations. The techniques developed in those efforts can be largely carried over to the task of giving structure/function representations of social systems of other sorts.

C. Task Analyses

The social system representation involves a representation of the various possible alternative histories of the social system, including the various components it could have and how it could go with these components, and their components etc. down to some limiting level of detail. A task analysis deals only with what "gets done" without reference to how it gets done (except insofar as the "how" is once again reduced to a new set of what "gets done") and is therefore formally a systematically incomplete process representation.

A structure/function representation of a social system will, therefore, lead directly to some number of distinct task analyses each specifying a set of outcomes which, if they were brought about, would constitute an instance of the functioning of the social system. Of direct relevance would be those task analyses which

involve tasks or "jobs" to be performed by an information system as a component of the social system.

The derivation of the IS requirements as a set of tasks performed by the IS as a component of the social system as given by a structure/function representation provides a conceptual validation of the task analysis. It therefore provides a rationale and justification for the selection of those requirements to place on the information system.

At the present time it appears that the set of specifications for a given information system is often significantly deficient in regard to a rationale showing (a) how those requirements are related, conceptually and empirically, to mission accomplishment, and (b) why the given set of requirements was selected over alternate sets. Deficiencies of this kind may be critical because if the system requirements are not well grounded in an adequate task analysis, then demonstrating that the information system meets the requirements does little to demonstrate the effectiveness of its contribution to the mission it was designed to serve.

D. Validation of Representation; Level 4 Evaluation

Social System representations may be validated by performing predictive simulation studies involving the prediction of any system histories which are both empirically accessible and substantively significant. Validation of SS representation provides a basis for two further evaluation efforts.

(a) The first is to provide empirical validation of the task analysis associated with the IS requirements.

(b) The second is to perform the special sorts of simulation studies involved in Level 4 evaluation as described in Section 4.2.4.

6.3.2 Empirical Generalization and Validation

It was noted above that even a conceptually valid task analysis of desirable social system histories leaves room, in principle, for undesirable histories to occur. This is because of the simplification involved in going from a process representation to a task analysis. Because of this, there is a significant need to show empirically that, in fact, whenever the performance requirements on the system components are met, the system functions successfully. Naturally, empirical evidence provides only some degree of assurance, not proof, in this regard.

A. Validation of Task Analyses

Although empirical validation follows naturally from the social system representation and the conceptually derived task analysis it is possible to do some limited empirical validation on existing systems. What is required is an explicit task analysis and the following sorts of data.

(1) performance data: does (or, to what extent does) the IS performance meet the specifications given by the task analysis?

(2) social system data: observation of the social system over some number of functional cycles provides a limited range of criterion data.

(3) role playing data: performance data may be augmented by role playing data (see illustration in Section 5.3.5).

B. Empirical Generalizations

The I & W network has a substantial number of component information systems. Given that reliable empirical generalizations reduce the load on individual evaluation efforts and that full scale evaluation is not always feasible, it seems highly advisable to begin to accumulate data systematically across individual evaluations so as to establish whatever generalities can be demonstrated empirically.

The establishment of empirical generalizations is, in a sense, an extension of or an elaboration of the empirical validation of the task analyses as described above. The empirical generalizations have to do with the relation of particular jobs or particular individual characteristics to any of the functional levels to which they contribute whereas the task analysis validation has to do with the entire set of tasks at a given level in relation to functioning at the next higher level (see Figure 7).

Data for empirical generalizations could be collected separately or as part of the evaluation process. The design for data collection and the analysis of the data should be done separately from the evaluation process, since there is no guarantee that exactly the same data would be suitable for both enterprises in every case (e.g., the evaluation procedure might be primarily precautionary whereas the generalization data would be affirmative).

6.3.3 Precautionary Evaluation

Perhaps the most direct and immediate application in I & W is the use of the Precaution Paradigm to review or augment existing evaluation or to design new evaluations.

A. Review and Augmentation

The Precaution Paradigm provides a rationale and a systematic way of obtaining assurances (or failing to obtain them) that the information system will not fail in the ways which are of greatest concern. Such assurances are in general important to obtain, since, as noted above, even conceptually and empirically validated task analyses will not entirely guarantee that the functioning of the information system will be free from serious failures.

The practical framework given by the Precaution Paradigm provides a basis for reviewing existing evaluations in order to establish whether the relevant assurances were obtained. If they were not, auxiliary procedures can generally be undertaken to obtain the relevant data.

Although existing evaluation efforts often demonstrate an intuitive grasp of precautionary thinking, they do not ordinarily provide consistent and effective implementation of precautionary considerations. For example, one extensive DOD evaluation effort now in progress was found to be essentially a set of precautions against a single eventuality; other eventualities and precautions meriting higher priority from a mission standpoint had been completely overlooked.

B. Evaluation Design

Particularly in conjunction with positive value demonstration methodology described above, the Precaution Paradigm is particularly significant for I & W applications. Even in the absence of conceptual and empirical task analysis validation it is often the case that an information system is introduced which has features which create the legitimate presumption that "in principle" or "on the whole" it ought to function successfully. Then the principal practical questions have to do with whether the system has flaws or limitations which would seriously limit or reduce its functional value. In such circumstances the precautionary approach can quickly get to the heart of the matter by addressing the relevant priorities and the corresponding data and assurances.

REFERENCES

- Anderson, R. H. "Rule-based decision aids and man-machine interfaces." Proceedings of the International Conference on Cybernetics and Society, Tokyo-Kyoto, Japan, November 3-7, 1978, pp. 1283-1288.
- Aronson, B. "Germany's war on noise." Mach. Des., 48, No. 9, p. 152, 1976.
- Ashton, R. G. "Simulation in the design and measurement of military command and control systems." Symposium on Human Operators and Simulation, Loughborough, Leics, pp. 117-122, March 1977.
- Atwood, R., & Kirschbaum, as summarized in Hackathorn, R. D., & Zeiger, H. P., "Interactive Dialogue Modes," 1979.
- Badler, N. I. "A concept model for the description of image sequences." Proceedings of the IEEE Milwaukee Symposium on Automatic Computation and Control, Milwaukee, Wisconsin, pp. 445-449, 1976.
- Baron, S. "The effects of deviate internal representations in the optimal model of the human operation." Proceedings of the 1976 IEEE Conference on Design and Control, pp. 1055-1057, 1976.
- Bell Telephone Labs, Inc. "Tactile man-machine communication system." U.S.P. 3919691, 1975.
- Bobrow, D. G., & Winograd, T. "An overview of KRL, a knowledge representation language." Cognitive Science, 1(1), pp. 3-46, 1977.
- Brosey, M. "Two experimental comparisons of relational and hierarchical database models." Int. J. Man-Mach Stud., 10, pp. 625-637, 1978.
- Brown, George R., Kornman, Brent D., Landauer, Christopher A., Mah, Clinton P., & Norris, Deborah J. "Message Extraction through Estimated Relevance (METER)." PAR Report, #79-77, 1979.
- Cardenas, A. F., & Seely, R. W. "A simple data-structure for interactive graphic design/drafting." Comput J., 18, No. 1, 1975.
- Carter, R. C. "Visual search with large color codes." Int Conf on Displays for Man-Mach Syst, Lancaster, England, pp. 65-68, 1977.
- Chambers, J. A., & Sprecher, J. W. "Computer Assisted Instruction: Current Trends and Critical Issues." Communications of the ACM, 23, No. 6, 1980.
- Codd, E. F. "A relational model of Data for Large Shared Data Banks." Communications of the ACM, 13, No. 6, pp. 377-387, 1970.
- Codd, E. F. "Normalized data base structure: A brief tutorial." Proceedings of the 1971 ACM/SIGFIDET Workshop, November 1971.

- Cooley, M. J. E. "CAD: A trade union viewpoint." CAD 76, JPC Science and Technology Press, Guildford, pp. 308-312, 1976.
- Corbett, C. "An interactive graphics system using an overlay tablet." Digital Processes, 2, No. 4, pp. 291-306, 1976.
- Davies, D. W., & Yates, D. M. "Human factors in display terminal procedures." Evolutions in Computer Communications, pp. 777-783, 1979.
- Denny, J. P. "Effects of anxiety and intelligence on concept formation." J. Exp. Psychol., 72, p. 596, 1966.
- Drake, S. K. "GRAPH--an interactive computer program used for technical plotting." Air Force Aero Propulsion Lab., AD-A086354/8, 101 p., 1980.
- Dunn, R. F. "Anxiety and verbal concept learning." J. Exp. Psychol., 76, p. 286, 1968.
- Elam, P. C. "Human considerations." Computerworld, March 31, 1980.
- Elliot-Green, R., & Parslow, R. D. Advanced Computer Graphics--Economics and Applications. Plenum Press, 1971.
- Encarncao, J. L., & Oldenbourg, R. Computer Graphics: Programming and Application of Graphical Systems. Munich: Verlag Combih, 1975.
- Enstrom, K. D. "Real-time determination of how a human has allocated his attention between control and monitoring tasks." IEEE Trans. Syst., Man & Cybern., Vol. SMC-7, No. 3, pp. 153-161, March 1977.
- Feigenbaum, E. A. "Artificial intelligence research: What is it? What has it achieved? Where is it going?" Symposium on Artificial Intelligence, Canberra, Australia, 1974.
- Feigenbaum, E. A. "The art of artificial intelligence--Themes and case studies of knowledge engineering." AFIPS Conf Proceedings, pp. 227-239, 1978.
- Feigenbaum, E. A., & Feldman, J. Computers and Thought. New York: McGraw-Hill, 1963.
- Fitter, M. "Towards more 'natural' interactive systems." Int. J. Man-Mach Stud., 11, 3, pp. 339-350, May 1979.
- Foley, J. D., & Wallace, V. L. "The art of natural graphic man-machine conversation." Proc. IEEE, 62, No. 4, pp. 462-471, 1974.
- Gaines, B. R., & Facey, P. V. "Some experience in interactive systems development and applications." Proceedings of the IEEE, 1975.
- Gedye, J. L. "A computer-based aid to self awareness." BIM Conference, 1972.

- Geise, D. M., Lubart, N. D., & Chang, I. F. "Resolution model for a CRT based person/machine interface." Proc. SID, 19, No. 3, pp. 105-111, 1978.
- General Accounting Office. "New ways of preparing data for computers could save money and time and reduce errors." PB-284286/2, 48 pp., 1978.
- Ghosh, S. P., & Senko, M. E. String path search procedures for data base systems." IBM Journal of Research and Development, 18,m No. 4, pp. 408-422, 1974.
- Goddard, P. J., Villalaz, P. A., & Spence, R. "Method for the efficient computation of the large-change sensitivity of linear non-reciprocal networks." Electron Letts., 7, pp. 112-113, 1971.
- Goldstein, D., & Papert, S. "Artificial intelligence, language, and the study of knowledge." Cognitive Science, 1(1), 1977.
- Greenblatt, D., & Waxman, J. "A study of three database query languages." In Sneiderman, B., ed., Databases: Improving Usability and Responsiveness. New York: Academic Press, pp. 77-97, 1978.
- Gregory, R. "On how so little information controls so much behavior." Business Research Report No. 1, Machine Intelligence Department, University of Edinburgh, 1968.
- Grover, D. "Hardware for visual information." Computer-Aided Design, 9(4), pp. 223-232, 1977.
- Guida, Giovanni, & Somalvico, Marco. "Interacting in natural language with artificial systems: The Donau Project." Inf. Syst., 5(4), pp. 333-334, 1980.
- Hackathorn, R. D., & Zeiger, H. P. User-Friendly Interfaces to Databases. Division of Information Science Research, Boulder, Colorado, August 1980.
- Harris, Larry R. "The Robot System: Natural language processing applied to data base query." ACM 0-89791-000-1, 1978.
- Hayes, P. "Trends in artificial intelligence." Int. J. Man-Machine Studies, 10, pp. 295-299, 1978.
- Herot, C. F. "A spatial graphic man-machine interface." Information Processing 80, pp. 1039-1044, 1980.
- Herscher, M. B., & Cox, R. B. "Operational real-time speech recognition systems." Proceedings of the 1975 Int. Conf. on Cybernetics and Society, pp. 300-301, 1975.
- Humcke, D. J., & Kent, D. P. "Ergonomics of a large interactive graphics operation." Computer-Aided Design, 9, pp. 262-266, 1977.

- Jandy, G. "Recognizing and utilizing heuristic rules in problem solving of operations research." Acta. Tech. Acad. Sci. Hung., 85(3-4), pp. 399-419, 1977.
- Jarvis, J. F. "A graphical display system utilizing plasma panels." Comput. and Graphics J., 1(2/3), 1975.
- Jeffrey, H. J., & Putman, A. O. The Mentor Project: Replicating the functioning of an organization. In K. E. Davis, (Ed.), Advances in Descriptive Psychology, Vol. 3. Greenwich, Conn.: JAI Press, in press, 1981.
- Jones, F. "Data processing in colour." Data Processing, 16(6), 1974.
- Jones, P. F. "Four principles of man-computer dialogue." Computer-Aided Design, 10(3), pp. 197-202, 1978.
- Keen, P. G. "Decision support systems and the marginal economics of effort." Working Paper No. 79-03-16, Department of Decision Sciences, The Wharton School, 1979.
- Kennedy, T. "The design of interactive procedures for man-machine communication." Int. J. Man-Machine Studies, 6, p. 309, 1974.
- Kennedy, T. "Some behavioral factors affecting the training of naive users of an interactive computer system." Int. J. Man-Machine Studies, 7, pp. 817-834, 1975.
- Kent, W. "Limitations of record-based information-models." ACM Transactions on Database Systems, 4(1), pp. 107-131, 1979.
- Kent, W. Splitting the conceptual schema. In Proceedings of the Sixth International Conference on Very Large Data Bases, 1980. (Available through ACM).
- Kleinman, D. L. "Some new control theoretic models for human operator display monitoring." IEEE Trans. Syst., Man & Cybern., SMC-7(11), pp. 778-784, 1977.
- Kutsch, J. A. "A talking computer terminal." Proc. 1977 Nat. Computer Conference, pp. 357-362, 1977.
- Laughery, Jr., K. R. "Human strategies in the control of processes." Proceedings of the Human Factors Society 21st Annual Meeting, San Francisco, 1977.
- Lees, R. "The man/display relationship." Int. Conf. on Displays for Man-Mach. Sys., Lancaster, England, pp. 73-74, 1977.
- Leonard, Jr., J. J. "Human performance validation of simulators: Theory and experimental verification." Proceedings of Human Factors Society 19th Annual Meeting, pp. 446-455, 1975.

- Levin, M. "An introduction to DIAM--Levels of abstraction in accessing information." ACM'78, Washington, D.C., December 1978.
- Levin, M., & Schneider, L. S. The Separation of Enterprise and Data Models. Database Research Project, Sterling Systems, Golden, Colorado.
- Llewelyn, A. I. "Special issue: CAD hardware." Computer-Aided Design, 9(4), p. 222, 1977.
- Lochousky, F. "Database management system user performance." Ph.D. dissertation, University of Toronto.
- Lochousky, F., & Tschritzis, D. "User performance considerations in DBMS selection." ACM-SIGMOD Proceedings, pp. 128-134, 1977.
- Logicon, Inc. "Design considerations for the man-machine relation's user interface to TAS." AD-A071207/5, 133 pp., 1979.
- Lomov, B. F., & Venda, V. F. "Human factors: Problems of adapting systems for the interaction of information to the individual: The theory of hybrid intelligence." Proceedings of the Human Factors Society 21st Annual Meeting, San Francisco, 1977.
- McCann, C. "Graphic display interaction. Part I: Literature review." AD-A055402/2, 14 pp., 1978.
- McLuer, D., Allen, W., & Weir, D. "The man/machine control interface-pre-cognitive and pursuit control." 1978 Joint Automatic Control Conference, October 20, 1978, Philadelphia, pp. 81-88.
- Mace, D. F., Harrison, P. C., & Seguin, E. L. "Prevention and remediation of human input errors in ADP operations." Inst. for Res., AD-A081730/4, 122 pp., 1979.
- Machover, C. "Computer graphics terminals: A backward look." Proc. of the SID, 15(1), 1974.
- Martin, J. Design of Man-Computer Dialogues. Englewood Cliffs, N.J.: Prentice-Hall, 1973.
- Martin, W. A. "Some comments on EQS, a near term natural language data base query system." ACM 0-8791-000-1, pp. 156-163, 1978.
- Maver, T. W. "Democracy in design decision making." CAD 76, IPC Science and Technology Press, Guildford, pp. 313-318, 1976.
- Meister, D. "The indeterminate man-machine system." Proceedings of the Human Factors Society 19th Annual Meeting, Dallas, pp. 1-3, 1975.
- Minsky, M. "A framework for representing knowledge." In Winston, P. H. (Ed.), The Psychology of Computer Vision. New York: McGraw-Hill, pp. 211-277, 1975.

- Negroponte, N. The Architecture Machine. Cambridge, Mass.: MIT Press, 1970.
- Newell, A. "Notes for a model of human performance in ZOG." AD-A049532/5, 30 pp., 1977.
- Newman, N. M. "A graphical technique for computer input." Comput. J., 11, pp. 63-64, 1968.
- Nickerson, R. S. "Man-computer interaction: A challenge for human factor research." IEEE Trans. Man-Machine Studies, MMS-10, p. 164, 1969.
- Nilsson, N. J. Principles of Artificial Intelligence. Palo Alto, Calif.: Koga Publishing Co., 1980.
- Oppenheim, D., Greenfield, A., & Hurst, E. G. "Window management systems users manual." AD-A089322/2, 25 pp., 1980.
- Ossorio, P. G. Classification Space Analysis (RADC-TDR-64-287). Rome, N.Y.: RADC, 1964.
- Ossorio, P. G., Dissemination Research (RADC-TDR-65-314). Rome, N.Y.: RADC, 1965.
- Ossorio, P. G. State of Affairs Systems: Theory and Techniques for Automatic Fact Analysis (RADC-TR-71-102). Rome, N.Y.: RADC, 1971.
- Ossorio, P. G. "What Actually Happens": The Representation of Real World Phenomena. Columbia, S.C.: University of South Carolina Press, 1978.
- Parsner, W. L. "The evolution and architecture of a high-speed workstation for interactive graphics." Proc. 1979 Nat. Computer Conference, pp. 165-173, 1979.
- Potts, J. "Computer graphics--whence and hence." Computer and Graphics J., 1(2/3), 1975.
- Prother, D. C., Berry, G. A., & Burmunday, J. M. "The effect of prompting and feedback on performance during learning stress and transfer of perceptual skill." Proc. Ann. Conf. Amer. Psychol. Assn., 7, p. 643, 1972.
- Quilliam, M. R. "Semantic memory," in Minsky, M. (Ed.), Semantic Information Processing. Cambridge, Mass.: MIT Press, pp. 216-270, 1968.
- Reddy, D. R. "Speech recognition by machine: A review." Proc. IEEE, 64(4), pp. 501-531, April 1976.
- Reingold, I. "Display devices--a perspective on status and availability." Proc. of the SID, 15(2), 1974.
- Richardson, J. J., Hackathorn, R. D., & Berger, H. P. Tutorial Assistance to Online Database Users. Division of Information Science Research, Boulder, Colorado, August 1980.

- Robertson, S., McCracken, D., & Newell, A. "The ZOG approach to man-machine communication." Carnegie-Mellon University, AD-A081088/7, 55 pp., 1979.
- Robertson, G., Newell, A., & Ramakrishna, K. "ZOG: A man-machine communication philosophy." AD-A049512/7, 58 pp., 1977.
- Rohlf, S. "Linguistic considerations for user interface design." Integrated Office Systems, pp. 189-196, 1980.
- Rosenthal, R. I. "The design of technological displays." Processing of Visible Language, 1, pp. 451-472, 1980.
- Rothsberger, H., & Rakotomalala, J. "F.I.R.S.T., a high resolution raster scan display system for interactive layout of text and figures." Eurographics, 80, pp. 291-302, 1980.
- Rouse, W. B. "Human-computer interaction in multi-task situations." IEEE Trans. Syst., Man & Cybern., SMC-7(5), pp. 384-392, 1977.
- Rumelhart, D. E., & Norman, D. A. "The active structural network," in Norman, D. A., & Rumelhart, D. E. (Eds.), Explorations in Cognition. San Francisco: W. H. Freeman, 1975.
- Sackman, H. "Experimental analysis of man-computer problem-solving." Human Factors, 12, pp. 187-201, 1970.
- Savoie, R. E., Brugher, J. S., & Bliss, J. C. "Development of a hand-held talk calculator for the blind." Proc. 1976 Nat. Computer Conference, pp. 221-225, 1976.
- Schank, R. Conceptual Information Processing. Amsterdam: North-Holland, 1975.
- Schank, R. C., & Abelson, R. P. Scripts, Plans, Goals, and Understanding. Hillsdale, N.J.: Lawrence Erlbaum Assoc., 1977.
- Schneider, L. S. "A relational view of the data independent accessing model." Proceedings of the 1976 ACM/SIGMOD Internatl. Conf. on the Mgmt. of Data, pp. 75-90, June 1976.
- Schneider, L. S. "A relational query compiler for distributed heterogeneous databases." SHARE 50 Proceedings, Denver, Colorado, March 1978.
- Schneider, L. S., & Connolly, T. W. "A generalized database system simulator based on the Diam I." Database Research Project, Sterling Systems, Golden, Colorado.
- Senko, M. E. "Diam II: The arnary infological level and its database language—Foral." DBM Research, Yorktown Hts., N.Y.
- Senko, M. E. "A data independent architectural model—Four levels of description." IBM Research Report RJ982, Yorktown Hts., N.Y.

- Senko, M. E., Altman, E. B., Astrahan, M. M., Fehder, P. L., & Wang, C. P. "A data independent architecture. Model I: Four levels of description from logical structures to physical search structures." IBM Research Report RJ982, February 1972.
- Shaket, E. "A model of man/machine communication in computer aided manipulation." Proceedings of the Int. Conf. on Cyber. & Soc., Washington, D.C., September 1977, p. 773.
- Shaw, M. L. "Conversational heuristics for eliciting shared understanding." Int. J. Man-Machine Stud., 11(5), pp. 621-634, 1979.
- Schneiderman, B. Software Psychology. Cambridge, Mass.: Winthrop Publishers, Inc., 1980.
- Spence, R. "Human factors in interactive graphics." Computer-Aided Design, 8(1), pp. 49-53, 1972.
- Spence, R., & Apperby, M. D. "On the use of interactive graphics in circuit design." Proc. IEEE Symposium on Circuits and Systems, New York, 1974.
- Stocker, F. R. "An investigation of optimization techniques for drawing computer graphics displays." Pennsylvania State University, N80-13814/2, 39 pp., 1979.
- Sutherland, I. E., & Johnson, R. "SKETCHPAD: A man-machine graphical communication system." Proc. SJCC, p. 329, 1963.
- Swain, A. D., & Guttman, H. E. "Human reliability analysis applied to nuclear power." Proceedings of the 1975 Annual Reliability and Maintainability Symposium, Washington, D.C., pp. 116-119, 1975.
- Thomas, J. C., Jr. "A design-interpretation analysis of natural English with applications to man-computer interactions." Int. J. Man-Machine Studies, 10(6), pp. 651-668, November 1980.
- Thompson, B. H., & Thompson, F. B. "Rapidly extendable natural language." ACM 0-89791-000-1, pp. 173-182, 1978.
- Topmiller, D. A., & Aume, N. M. "Computer-graphic design for human performance." Proc. Annual Reliability and Maintainability Symposium, Los Angeles, pp. 385-388, 1978.
- Vaccarino, C. A. "Real-time simulation for verification of tactical electronic warfare systems." Communications of ACM, 0-89791-000-1, pp. 334-341, 1978.
- Vallee, J. "There ain't no user science: A tongue in cheek discussion of interactive systems." Proceedings of the ASIS Annual Meeting, 1976, p. 36, 1976.

- van Gigch, J. P. "A methodological comparison of the science, systems and metaphysical paradigms." Int. J. Man-Machine Studies, 11, pp. 651-663, 1979.
- Walsh, B. "Improving the man-machine interface." Electron. Electro-Optic Infrared Countermeasurer, Vol. 2, No. 6, pp. 53-55, 1976.
- Welch, J. R. "Automatic data entry analysis." AD A0459396, 141 pp., 1977.
- Welt, M. J. "NOMAD: A printed wiring board layout system." Proceedings 12th Design Automation Conference, p. 152, 1975.
- Wieselman, D. L. "Hard-copy computer output and its future." Proc. 1977 Nat. Computer Conference, pp. 363-370, 1977.
- Williges, R. C. "Experimental designs for investigation of complex human operator/machine systems." Proceedings of the Human Factor Society 21st Annual Meeting, San Francisco, 1977.
- Winograd, T. Understanding Natural Language. Academic Press, 1972.
- Winston, P. H. Artificial Intelligence. Reading, Mass.: Addison-Wesley, 1977.
- Witten, D. H. "Software for speech outputs." Information Processing, pp. 297-301, 1977.
- Witten, H., & Griffith, P. "A dual-processor structure for high quality real-time linear predictive analysis of speech." Digital Processes, 5(3-4), pp. 257-269, 1979.
- Woods, W. A., et al. "The lunar sciences natural language information system." BB&N Report 2378, June 1972.
- Zeiger, H. P., & Hackathorn, R. P. Conceptual Design for a Materials Database. Division of Information Science Research, Boulder, Colorado, July 1980.
- Zimmerman, R. "The design of systems for communications between men and machines." Regelungstech Prax., 18(2), pp. 45-49, 1976.

At the present time there is a voluminous literature dealing with evaluation of various sorts. The basic methodology consists of the classic experimental design models of behavioral science (Analysis of Variance and its linear additive variations such as multiple regression, factor analysis, correlation, etc.) augmented by some "quasi-experimental" designs which offer partial approximations to classic experimental designs. Certain difficulties arise when this methodology is applied to the evaluation of information systems.

1. Operationalization

Evaluative criteria are necessarily human oriented because information systems are designed to play a part in human activities and further human ends. In general, these criteria cannot be effectively operationalized in conventional experimental ways. For example, two familiar criteria are "ease of use" and "appropriateness of data," neither of which can be rigorously reduced to anything which is readily observable or countable and which can thereby serve as a conventional operationalization of these criteria. Efforts to make this kind of substitution typically meet with some agreement, some objection, and some outright rejection. (The nature of this difficulty is elaborated further in the discussion of empirical identities in Appendix D.)

2. Combinatorial Problems

Information systems have a multiplicity of aspects relevant to evaluation, most of which involve the criterion problem noted above. The fact that there are multiple aspects creates a new problem, namely, how to combine information or evaluations of various aspects into a single evaluative measure or criterion. Typically, the choice has been between (a) refusing to provide an overall evaluation, (b) combining partial evaluations in a more or less arbitrary

way, or (c) abandoning an analytic approach in favor of a global evaluation, with its attendant criterion problems. (A further discussion of the combinatorial or "tradeoff" problem appears in Appendix F.)

3. Relativity

The value of a given system configuration is relative to some human purpose, goal, or standard, and usually there is more than one of these in the picture. Unless these factors are stipulated or universalized (un-relativized) no single evaluative measure or characterization will in general be available. On the one hand this leads to the combinatorial problem again, since one could arbitrarily combine evaluations relative to several purposes and emerge with a single measure or criterion. On the other hand, it violates both a human tendency to suppose that there is a single characterization of "how good it is" and some informal, but strong, experimental conventions to the effect that only absolute data but not relative data are "objective." (Issues of relativity and context are dealt with further in Appendix D.)

4. Implementability

There is a very serious question as to whether an analysis of variance (ANOVA) approach can actually be carried out on complexly determined phenomena such as social systems or information systems (or an egg beater, for that matter). It is characteristic of this approach that

- (a) it generates an N-dimensional matrix of data cells (N variables or parameters each having some number of distinguished values);
- (b) data within one cell are independent of data in other cells and therefore give no information about other cells;

- (c) hence all cells in the matrix must be independently filled with data. (In the present context the within cell data would be evaluative data and the N variables or parameters would be the factors which are candidates for influencing the evaluative data.)

It is generally recognized that for complex phenomena the number of cells is impossibly large. For example, a mere three parameters each having ten values will generate 1000 cells. If the experimental effort required to obtain the data for a single cell is at all substantial, filling in 1000 cells is out of the question. (For an information system the number of cells would be on the order of 10^6 - 10^9). Williges (1977) suggests a variety of palliatives for this problem, e.g., single observations per cell, systematically incomplete sampling of cells, and so on, but they are palliatives, not solutions, and all involve a degradation in the confidence and generalizability of the results.

6. Appropriateness of Model

Analysis of Variance is most effective when there is a small number of independent variables, when these variables operate independently of one another, and when the full range of values of the independent variables is available for manipulation by an experimenter. None of these conditions hold for complex, structured processes such as information system or social system functioning.

- (a) As noted above, the number of independent variables is not small.
- (b) In configurational phenomena, e.g., complex social or physical machinery, influences tend to be totally interdependent rather than independent. In ANOVA terms this results in approximately all interaction effects being significant. But significant interaction effects strongly limit what one

can say about the influence of the independent variables. (See curve-fitting alternative below.)

- (c) For a real world phenomenon such as the operation of an information system, an experimenter is almost completely powerless to generate at will all the different ways that the system could function (i.e., all the values of the independent variables in combination). Yet some of the possibilities which are in this way beyond experimental observation may be most significant in determining the value of the information system (e.g., functioning under wartime conditions).

Analysis of variance methodology is designed to provide a warrant for statements to the effect that a dependent variable has different values as a function of the difference in the values of one (main effects) or more (interaction effects) independent variables. It is not designed to elicit functional dependencies. For the latter purpose we often turn to curve-fitting. However, there are some strong limitations in this regard.

- (a) Not all types of curve can be fitted using analytic mathematical methods.
- (b) Fitting N different types of curve simultaneously is beyond the state of the art except for very simple types (e.g., as in factor analysis).
- (c) There is no reason to suppose that the mapping functions applicable to information system variables and evaluative dimensions will be mathematical in nature.

From these considerations it appears that an appropriate methodology will involve at least the following.

- (a) Semantic (real world) modeling capability rather than merely quantitative characterizations.

- (b) Concept driven data gathering rather than data driven summary description.
- (c) Formal capability for handling discontinuous, irregular, and nonquantitative functional relations.
- (d) Formal capability for extrapolation to anticipated but unobserved situations.

These features are provided by the Precaution Paradigm and the Simulation Paradigm summarized in Appendix B and referred to in Appendix E.

APPENDIX B

**MATERIAL ON THE PRECAUTION PARADIGM, THE SIMULATION
PARADIGM, AND ACTOR-OBSERVER-CRITIC**

(Reproduced by permission of
Linguistic Research Institute)

Linguistic Research Institute

Report No. 22A

REPRESENTATION, EVALUATION, AND RESEARCH

Peter G. Ossorio

1979

© Copyright 1979
All Rights Reserved

Linguistic Research Institute
Boulder, Colorado
Whittier, California

In What Actually Happens I made the point that a new beginning was needed if we were ever to achieve a behavioral science. I also made the point that such an enterprise called for the assured compatibility of explanation, methodology, and historical description, and that the only way to assure that compatibility was to formulate a single conceptual framework for all three.

Given that one of the principal dogmas of la dolce vita academica is the categorical separation of method and theory, it is not surprising that the primary reactions among academicians and experimental practitioners range from blank incomprehension to incredulity to simple rage and lashing out. Even the two prima facie examples in What Actually Happens, where a substantive problem was dealt with by recourse to a 'methodological' formulation and a methodological problem was dealt with decisively by recourse to a 'substantive' formulation, have drawn little more than efforts to explain them away or dismiss them as being very special cases rather than exemplars of a general possibility.

The present paper involves a third and fourth example of this general possibility. In both cases, methodological formulations (research paradigms) are derived from 'substantive' formulations. In point of fact, as I mentioned in WAH, it is in some ways misleading to speak only of crossing over from substantive to methodological and vice versa. In most cases it would be apt to say that the conceptual formulations in question can be applied equally to either substantive or methodological issues. We speak of crossing over primarily when

the requirement of compatibility is relevant.

Although this paper has that polemic aspect, it is not merely a polemic exercise. If anything were needed to convince us that it is not merely in principle, but in a simple practical sense as well, that we need an explicit, systematic, coherent development of the rationale for scientific practice, we have only to take a hardheaded look at the situation in dolce academica, where the bureaucratic and political structure of 'behavioral science' has almost completely obscured and subverted its original intellectual aspirations. What we find there is, except in detail, what we might expect in any large and established bureaucracy, namely, ad hoc conventions, folklore, customs, territorialities, authoritative pronouncements, cliques, self-perpetuation, log-rolling, and a hypertrophied development of administrative procedures, which in this case are codified in mathematical and quasi-mathematical models for experimentation and statistical analysis. What we do not find, though there are islands of rationality here and there, is an intellectually responsible, coherent, and substantively adequate body of thought concerning the rationale and technology of the scientific study of behavior.

One of the advantages of making a fresh start is that we need not be forever enmeshed in fruitless dialogues whose essential failure is guaranteed by the inadequacy of the general intellectual framework administered by dolce academica (the "Tar Baby problem" in WAH). However, two simple examples of what one might take issue with may be informative.

(a) The convention of 5 percent and 1 percent significance levels is a fairly obvious case of bureaucratic convenience taking

precedence over task-oriented rationality. The pretence that 5 percent significance is decisive evidence but 5.01 percent significance is no evidence at all has a special kind of absurdity. And the ad hoc conventions associated with it lead to further absurdity. For example, I may have three groups A, B, and C, with a significant difference between A and C but non-significant differences A-B and B-C; under these conditions I am required to conclude that A is different from C, but B is not different from either A or C. This absurdity results from what amounts to substituting my degree of certainty about the differences for my quantitative estimate of what the differences are. And although this procedure might well be formally repudiated when stated so baldly, in academica it is enforced administratively with a heavy hand by editors and editorial reviewers in actual practice. In the face of such conventions, one might well feel the need to come back through the Looking Glass to the real world.

(b) Much of what is now called "methodology" and "experimental design" is based on the idea that one samples from a population and generalizes one's results to other samples or individual members of that population. Yet it is almost never possible, even in a practical sense, much less in a methodologically rigorous way, to specify what population the subjects in a psychological experiment are to be taken as a sample from. The corresponding difficulty arises in justifying one's generalization of the results. (These issues can in general be paraphrased as issues concerning the representativeness of a sample with respect to a "population.") The convention in dolce academica has been to give an elliptic description of one's sample

as the description of the population one generalizes to or, more prudentially, to say nothing. For example, if my subjects are a particular cohort of sophomores in psychology classes in a particular institution, I will gravely announce that I generalize my results to "college sophomores" or "college sophomores at this institution." But nobody believes for a moment that my subjects are representative of such 'populations.' In short, as a member of dolce academica I must generalize, for experiments are not yet formally recognized as ends in themselves, but I cannot do so legitimately. After almost one hundred years of psychological research we have introduced the neologisms "internal validity" and "external validity" so as to limit the scope of the problem to the latter. Somehow, the introduction of these terms has not really solved the problem of generalization. It has, perhaps, made it easier to recognize the enormous difference between the existing administrative regulations concerning what qualifies as doing an experiment properly and the genuine methodological requirements for effective research concerning people and their behavior. (For an enormously different approach to generalization, see "Explanation, Falsifiability, and Rule-following.")

What we shall require, and what I propose to accomplish in a limited way here, is an explicit, systematic, and coherent development of the methodology of scientific practice in the study of persons and their behavior. Since that scientific practice encompasses an organized variety of behavioral forms, our task is to represent those forms normatively; hence there is the possibility of deriving our representations as special cases from more general formulations of

persons and their behavior. The specific task for an explicit methodology is to formulate as clearly as possible the structures of behavior and decision which are relevant for investigating persons and their behavior and formulating (and justifying) our understanding thereof. We will deal with two such structures here. They are in addition to the Pragmatic Paradigm presented in Persons and the Demonstration Research Paradigm presented in "Explanation, Falsifiability, and Rule-following."

The point of this is not to provide pre-packaged answers which are to be accepted without question nor to present procedural prescriptions which are to be implemented without question. On the contrary, it is to facilitate both (a) a descriptive appraisal of whether a given set of behaviors exemplifies a given methodological structure, and (b) a critical appraisal of whether the procedures which fit a given structure are relevant or appropriate to the task or question at hand on a given occasion. If behavioral science is to be a rational form of behavior, such critical appraisal cannot be abdicated in favor of political-administrative fiats and clerical-mathematical procedures.

In this respect I propose to (a) review two conceptual schemas which are used in Descriptive Psychology to represent real world phenomena, (b) delineate two general research paradigms which exemplify these conceptual schemas, and (c) illustrate these research paradigms with some special reference to evaluation research. I have chosen evaluation research because it appears that, more, even, than behavioral research generally, this field is afflicted by scientific

posturing, shamanism, and politics as usual.

Although I referred earlier to making a fresh start, when it comes to evaluation research it would be more accurate to speak of having made an early start. Much of the Descriptive rationale for evaluation research was developed in the course of a research program (Ossorio, 1964, 1966a, 1966b, 1967, 1968, 1971) which began in 1961, long before the emergence of Evaluation Research as a distinctive social enterprise. It would be surprising if the later developments in that field bore no resemblance to the developments in Descriptive Psychology. Such similarities are not to the point here. I am not concerned with the fragmentary ideas and techniques which have emerged here and there in the Evaluation Research literature, nor is the present report designed to contribute to that literature. I am concerned with delineating a comprehensive and coherent methodology which has evolved quite independently within Descriptive Psychology.

I. The Actor-Observer-Critic (AOC) Schema

This schema distinguishes three methodologically fundamental behavioral statuses, or ways of functioning. They are characterized individually as follows.

A. In functioning as Actor, the individual acts in response to his circumstances in accordance with his nature. This form of functioning is, therefore, characterized as (1) before-the-fact, (2) spontaneous, (3) creative, and (4) value-giving. It should be noted that although Actor, Observer, and Critic are distinguished as methodological roles rather than full-fledged human roles, there is

a natural tendency to think in the latter terms. In this case it is helpful to think of Actor functioning as involving both observation and appraisal but of a purely assimilative sort. That is, as Actor, one notices and makes use of things only insofar as they fit into, or affect, one's ongoing project. It is their place in one's life and activities which gives value to those things, hence the characterization "value-giving."

B. In functioning as Observer-describer, the individual takes note of (discriminates, describes, distinguishes) the course of events. Although particular attention may be paid to those things which are relevant to the behavior of the Actor, this form of functioning is objective, i.e. normative, rather than egocentric or merely subjective, as in Actor functioning. It is therefore characterized as (1) after-the-fact, (2) passive, (3) reflective, and (4) either value-neutral or value-finding.

C. In functioning as Critic-appraiser, the individual begins with the results of Observer-describer functioning and makes an appraisal of whether the course of events is satisfactory or unsatisfactory. If satisfactory, the Critic takes no positive action but appreciates/enjoys the success. If unsatisfactory, the Critic generates (a) a diagnosis, i.e. an account of what it is that has gone wrong and/or (b) a prescription, i.e. a practical guide in regard to what to do differently so as to improve matters. This form of functioning is, accordingly, characterized as (1) after-the-fact, (2) reflective, and (3) judgmental, verdict-giving, or value-finding.

There are two forms of the AOC schema. In the "methodological form," it has been used, among other things, to generate a criterion of adequacy for general theories of behavior (Ossorio, 1973). In this connection it may be noted that Critic-appraiser functioning is a special case of Observer-describer functioning, and the latter is a special case of Actor functioning.

In the "functional form" of the schema, the three statuses are related in the form of a feedback loop. As shown in Figure 1, what the critic judges to be the case is an essential part of the circumstances to which the actor then responds.

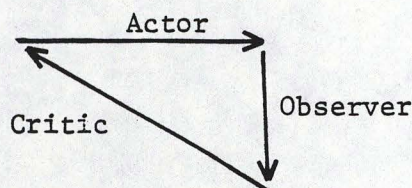


Figure 1. Functional AOC Schema

Formulas and schemas are, of course, not descriptions. The schematic differentiation of these ways of functioning is not a statement to the effect that there are different processes going on or that the person is engaged in three behaviors simultaneously, etc. In point of fact, a person is normally functioning in all three ways simultaneously. In any continuously functioning feedback loop, all the elements are operating simultaneously, yet the feedback function depends on a temporal succession and coordination among the three essential elements.

One of the primary values of the Actor-Observer-Critic schema is that it provides a systematic representation (in conjunction with

other conceptual-notational devices) of a person as being self-regulating while nevertheless being responsive to both personal and circumstantial states of affairs. Unlike common 'psychological' approaches such as Attribution Theory or Operant Conditioning Theory, which portray self-influencing mechanisms, the AOC schema permits us to portray the kind of self-regulation which is the primary expression of human rationality, though, of course, it does not force us to do so. It is this feature which finds a fairly direct exemplification in one of the research schemas, the Precaution Paradigm. In this paradigm, the collection of data is a case of rational precaution-taking or assurance-seeking. The methodology in this case has to do with the rationale of the precaution-taking; the technology, if there is one, has to do with the implementation (operationalization, realization) of the precaution-taking. In neither case is mathematical modeling (e.g. sampling theory, statistical analysis, measurement) or standard 'experimental design' necessarily involved.

II. The Precaution Paradigm

Some version of the classical experimental designs involving a control group and experimenter manipulation of independent variables is commonly considered to be the standard of rigor in establishing general conclusions on an empirical basis. Correspondingly, "applied research," including validation or evaluational research, is generally seen in either one of two lights. (a) In the first case, it is seen as merely the application of general "knowledge" which was itself acquired in this foolproof (rigorous) way. That is,

it is seen as a mere demonstration that some general principles arrived at via 'rigorous' methodology do apply to some phenomenon of interest in the real world. (But recall the problem of generalization and "external validity.") (b) In the second case the classic experimental procedures are held up as the ideal to be striven for in the empirical examination of the real world phenomenon itself. It follows in this case that to the extent that the classic design is not adhered to (and usually, it cannot be) the results are of dubious value and the conclusions drawn are unsound (though possibly true).

It should be noted however, that classical experimental methods are designed to support general conclusions. The primary interest is never on the actual subjects as such. (This is why the illegitimate generalization is required.) In contrast, much 'applied' research, and certainly most evaluation research, is designed to support particular conclusions. That is, what is of primary interest is not whether a certain kind of program, procedure, instrument, etc. is generally successful in this or that way or degree, but rather whether a historically particular program (or set of programs), procedures, etc. is in fact successful in this or that way or degree over a particular span of time.

Such a statement points up the contrast between the methodological structure of the precaution paradigm and the intuitive judgment of the traditionalist. For, in accordance with (b) above, the latter would be very likely to object. Thus:

Wil: That just shows how all this philosophizing can lead you astray. Of course all investigations aim at some degree

of generality. Just because you're not talking about a real experiment, that doesn't mean evaluation doesn't aim at some generality. Even a particular evaluation has to assume that if a program is effective it would also be effective with a similar population in similar circumstances. Otherwise, you'd just be engaging in an empty exercise of assigning praise or blame. Surely you don't want to do that.

Gil: You do have a penchant for playing "Now I've Got You, You Son of a Bitch," don't you. First off, let me say that what I was describing was evaluation research, not "an investigation." Secondly, let me tell you flatly that when I, as an administrator, commission a job of evaluation research or do it myself, what I want to know is how well this program is doing, and I don't give a tinker's dam about some other program, and I sure as hell don't care what someone might say about some purely hypothetical 'similar' program in some purely hypothetical 'similar' circumstances with some purely hypothetical 'similar' populations. I deal with the real world, not with mystic verbal or Pythagorean formulas.

Thirdly, I can tell you just as flatly that, whether or not the evaluation involves praise or blame, the basic fact is that I'm going to act on that appraisal. I'm going to change the program, leave it alone, or axe it. You'd better believe that's not a meaningless exercise. By the way, tell me again about how you guarantee the external validity of an experiment and how you know what to generalize to. If you want a meaningless exercise, I give you that one for free.

The dispute and polemics are themselves not merely ad hoc. They, too, can be clarified by reference to the AOC schema. The traditionalist stance clearly reflects a combination of Observer-describer functioning and the "diagnostic," or verdict-rendering aspect of Critic functioning. Thus, Wil thinks primarily of judgmental descriptions as the end product and worries about their generality. Just as clearly, the precaution paradigm codifies a stance that reflects primarily Actor functioning and the "prescriptive" aspect of Critic functioning. Thus, Gil regards his later actions as the primary product of the evaluation and worries about their justification.

Even when more general conclusions are desirable in evaluation research, the emphasis is properly on a case by case approach because ordinarily the relative influence of historical-situational contexts is so marked that simple generalizations about procedures (programs, etc.) across contexts would be rash or foolhardy (even if it were a case of 'applying rigorously established findings'). One can always sum across replications of a fine-grained analysis; one cannot recover individual patterns from group data.

Clearly, then, the specifications for a viable evaluation methodology would include (a) being capable of rigorous implementation in the real world setting, (b) providing fine-grained or single case conclusions, and (c) lending itself to systematic variation and replication for supporting more general conclusions. The Precaution Paradigm is responsive to these requirements.

Let us first examine the fine-grained module. In this case we presuppose a real world setting where there is a procedure (instrument, program, etc.) which is engaged in and which is a candidate for evaluation. Our first move is to construct a representational formula (cf. the Process Description below) for the phenomena in question. This will have a general form of the following sort: A uses/does B with C in ways D (e.g. to accomplish G), which can go wrong in ways E_i , as indicated by observations F_{ij} .

For example, a therapist (A) may use a projective test (B) with a client (C) by interpreting M responses (D) to arrive at a psychodiagnosis and differential treatment (G). This might go wrong in that the client was misdiagnosed and the treatment would be ineffective (E_i), as indicated by the continued presence of the initial symptoms (F_{ij}). Or again, an agency worker (A) may use a "mastery of English" entry on an application form (B) with an unemployed refugee (C) to direct him to a laborer's job and a vocational training class (D). This might go wrong in that the refugee would perceive himself as having lost face and subsequently become depressed (E_i), as indicated by self-report of feelings and symptoms and failure to hold the job or progress in training (F_{ij}).

Given the representational formula, we next focus on the ways in which A could go wrong by using/doing B with C in those ways. At this point there arises the issue of real, or practical, doubt versus idle skepticism. In this regard, we introduce two test questions. The first is, which of these ways of going wrong am I most worried about? Given an answer to that, the second question is, is that possibility important enough so that it's worth taking precautions

against going wrong in that way? If the answer here is "yes," the next question is, which of the indicators of its going wrong that way are both convincing enough and accessible enough to make them worth getting (via the relevant observations)? If there are such indicators, the observations are made. These observations provide either a reassurance that the project is not going wrong that way or a warning that it probably is going wrong in that way. One of the implications of speaking of "indicators" rather than "measures" here is that neither measurement nor any quantitative procedure is essential here, although both are allowed for.

The foregoing constitutes the required methodological unit or module. In this regard the following may be noted.

- (a) The sequence, and the logic of the sequence, is that of Actor-Observer-Critic, and the sequence is one of an error-detecting feedback cycle.
- (b) What qualifies as "going wrong" (the "diagnosis") is context-bound insofar as it depends on the specific purpose(s) for which B is used/done and the specific setting, persons, et cetera which are involved.
- (c) Correspondingly, the indicated corrective measures (the "prescription") will be context bound. (In the context of program or treatment evaluation, an overall evaluation will correspond to a single feedback cycle whereas a pragmatic evaluation will involve repeated feedback cycles and modification of the program during its progress if the feedback information so indicates.)
- (d) The functional AOC schema is both an error-detecting and a success-detecting feedback loop. Although one might define success logically

in this context as simply the absence of failure, in fact there may be, and usually are, positive indicators of success. Thus, there is a corresponding module based on a modified formula in which we refer to ". . . can go right in ways E_i , as indicated by observations F_{ij} ." And, of course, we may combine the two, i.e., ". . . can go right or wrong in ways E_i etc."

- (e) What qualifies as success will in general be no less specific and context-bound than what qualifies as failure.
- (f) If there are no observational indicators of B having gone wrong in a given way in the real world settings, we may conclude that there is no real problem of its going wrong in that way.

The methodological module lends itself to systematic variation and replication because the representational formula amounts to a parametric analysis of the kind of use (instrument, program, etc.) which is to be evaluated. That is, A, B, C, D, E, and F will be parameters of that kind of use. Thus, one may replicate across A, i.e., persons or groups engaging in the procedure; across B, i.e., different forms, instances, or variations of the procedure; across C, i.e., different recipients or groups of recipients of the procedure; across D, i.e., the various ways of using/doing B, with special reference to the purposes, circumstances, settings, etc.; across E, i.e., ways of going right or wrong; and across F, i.e., different indicators of success or failure. And, of course, one may replicate across any or all combinations of these various possibilities.

By way of elaboration we may note the following.

- (a) Although the paradigm presented above involves the representation of a procedure in a historical setting, the procedure may be either actual or hypothetical, e.g., merely intended or planned. Correspond-

ingly, and particularly in the latter case, the ways in which the procedure might go right or wrong might be exemplified in some other setting, e.g., an experimental setting. For example, if the failure envisioned in the historical setting were the misdiagnosis of the client, that failure might be equally well exemplified in an experimental setting with clients selected specifically for the purpose of conducting the test. (But it also might not be. The problem of the "external validity" of experiments is a real one and there are no general solutions to it.)

- (b) The modular approach makes relatively clear something which is often glossed over in the classic experimental tradition, namely that it is impossible to take all logically possible precautions against being wrong or to obtain all logically possible assurances of being right. I say "glossed over" not in the sense that anyone would deny it overtly, but rather in that (a) taking more precautions is fairly automatically counted as being "more rigorous" than taking fewer precautions, and (b) pointing out a possible precaution which in fact was not taken is almost automatically counted as a legitimate criticism. These actions on the part of experimental and editorial practitioners speak louder than their words. These actions are compatible with the principle that one can and should take all possible precautions. They are not compatible with the principle that precautions reflect a critical appraisal and must be justified in each case.
- (c) Classic experimental design can be seen as a case of (1) taking certain standard precautions, whether relevant and important or not, and (2) taking them in advance of serious real world practice

('application'). Such a strategy might have some real value, but it might also be counter-productive, since the inability to meet the a priori requirement of certain precautions is likely to result in (a) the failure to do needed research or (b) the stigmatizing of such research as unsuccessful.

- (d) Experimental procedures, like any procedures, will exemplify B in Formula (1). We may represent the real world phenomenon by saying that A uses experimental procedures (B) with subjects and problems (C) in certain ways and with certain purposes in particular settings (D) such that A can go wrong in ways E_i as indicated by observations F_{ij} . This holds equally for the use of traditional experimental paradigms and for the newer paradigms such as the Precaution Paradigm.

The use of experimental procedures is thus not something which per se carries any guarantee of any kind of success or avoidance of failure. And we cannot, with respect to a given kind of use of experimental procedures take all possible precautions etc. Experimental procedures provide a framework for exercising human judgment and competence, not a way of doing without it.

This conclusion may appear to be truistic. True. In classic 'methodology' it is merely an ad hoc, commonsense truism, not part of the methodology. In Descriptive Psychology, it is a methodological truism. In classic 'methodology' the necessity for exercising judgment in research appears extraneously in the form of exhortations to be knowledgeable or to be careful in this or that way. In Descriptive Psychology that necessity is an explicit part of the general behavioral concepts of which methodological concepts are special cases, and so is in no way extraneous. But then, classic

'methodology' is essentially only an elaborated technology, so such differences are to be expected.

III. Process Representation

The systematization of reality concepts (object, process, event, state of affairs, relationship) via a transformational calculational system and the systematic explicit representation of objects, processes, et cetera are presented in "What Actually Happen" (Ossorio, 1975, 1978). The following is a brief summary and review.

The two most relevant transformational rules ("Transition Rules") dealing with the general concept of a process are as follows.

Rule 4. A process is a sequential change from one state of affairs to another.

Rule 5. A process is a state of affairs having other, related, processes as immediate constituents.

Because these recursive rules deal with the general concept of a process they apply to all processes and will therefore not distinguish one process from another. In order to have a general method for giving explicit representation to a particular process or kind of process we require (a) a parametric analysis of "process" and (b) a systematic notation or format for giving process representations. Both of these requirements are met by the "Basic Process Unit" (BPU) shown in Table 1.

The Basic Process Unit has a gross structure of "Name" and "Description." The former identifies the process and the latter gives the explicit representation.

In the explicit "Description" we may distinguish a "gross structure"

(Stages, Options) and a "fine structure," or "State of Affairs Structure," (Elements, Eligibilities, Individuals, Contingencies, and Versions).

Rules 4 and 5 are most clearly expressed in the fact that the process is represented as a sequence of Stages each of which is itself a process. The fact that the latter processes (stages) can each have some number of distinguishable exemplars is expressed in the association of some number of Options (exemplars) to each Stage. It is the Stage-Option structure which codifies the recursiveness of Transition Rules 4 and 5, for each Option is itself a process and can therefore be represented by a Basic Process unit involving a new set of Stages and Options, and the latter can in turn be so represented, and so on ad infinitum. This enables us to represent processes of any magnitude in any degree of detail, down to continuous processes.

Processes generally involve object constituents as well as process constituents. These objects have certain relationships (their having these relationships is a state of affairs) which change over time and the changing of these relationships over time is (the same state of affairs as) the occurring of the process. "Element" refers to the formal ingredients (objects) of the process. "Individuals" refers to historical individuals in the abstract, and "Eligibilities" assigns Individuals to Elements. For example in the play "Hamlet," Hamlet, Polonius, the skull, and the castle are all Elements. Person X, Person Y, prop A, and prop B are abstract historical individuals. Either Person X or Person Y may be Eligible to play Hamlet or Polonius (but not the skull or the castle), whereas Prop A is eligible to play the part of the skull and Prop B is eligible to play the part of the castle. Both Individuals and Elements are needed in the formal specification because there need not be a one to one relation between them. For example, Person Y may

be eligible to play Hamlet in Scene I and to play a spearbearer in Scene II. Actual performances of "Hamlet" require actual individuals in place of the abstract Individual even though "Hamlet" as a kind of process involves no reference to particular persons.

The occurrence of a process involves the occurrence of one of the Options for Stage 1 followed by the occurrence of one of the Options for Stage 2, and so on. In the process representation restrictions on the occurrence of particular Options are given by Contingency statements. Contingencies may be Attributional or Co-occurrence contingencies.

In the case of a Co-occurrence contingency the occurrence of a given Option in a given Stage is incompatible with (or necessitated by) the occurrence (or non-occurrence) of some other Option in some other Stage (or a combination of such options). For example, in a chess game, White's possible fifth moves are highly restricted by which of the possible first, second, third, and fourth moves have actually occurred, and many moves which are in principle possible in chess as fifth moves (are Options in Stage 5 of that process) are not possible in this game (this Version (see below) of that process).

In the case of an Attributional contingency the availability of a given Option in a given stage is contingent on some attribute of the individual who is serving as a given Element. For example, throwing a 90-yard pass in football is one of the formal possibilities, but it could only occur if the passer had an exceptionally strong arm.

Statistical Contingencies represent empirical correspondences rather than conceptual requirements. They may be of either a Co-occurrence or Attributional sort. For example, a 90-yard pass would be unlikely if the line of scrimmage were more than 30 yards from the offensive team's goal

line (which reduces to a Co-occurrence Contingency). Similarly, a checkmate in four moves would be unlikely to be available to a player who did not see it ahead of time. And a refugee for whom attending class represented an extreme loss of face would be unlikely to attend class.

When a process takes place it takes place in one of the ways in which it can take place. The specification of Stages, Options, Elements, Individuals, Eligibilities, and Contingencies is a way of specifying the conceptual restrictions (optionally, augmented by empirical restrictions) on what sort and sequence of happenings would qualify as an occurrence of (an exemplar of) the process in question. Each such distinguishable exemplar is a Version of the process. Each of the ways in which a given process can take place is a Version of that process. The occurrence of the process on a given occasion is (the same thing as) the occurrence of one of its Versions on that occasion. In contrast, occurrences of the same process on different occasions will generally involve the occurrence of different Versions on the different occasions. And different Versions need not resemble each other in any other way than in being Versions of the same process.

The problem of "generalizing" or "applying" the results of laboratory or 'analogue' research is in part the problem of trying to draw conclusions about one process (Version A of Process X) on the basis of observing another process (Version B of Process X) which is not unlikely to be different in many crucial respects.

The requirements for describing a given process as such are in principle not different from the requirements for a parametric analysis of "process," i.e., for specifying the ways in which one process (or kind of process) can be the same as another process (or kind) as such or different from it. Thus, the reference to Stages, Options, Elements, Individuals,

Eligibilities, Contingencies, and Versions constitutes a parametric analysis of "process."

One important feature of the process representation provided by the BPU format is that it exemplifies a holistic, or "top down," approach. Description consists of specifying various facts about the process in question. Each fact serves to further distinguish that process from other processes of the same general sort, and there is no definitive 'complete' description. Thus, the BPU form of representation begins to be informative as soon as any information is available concerning the process in question.

Systematically incomplete process descriptions are codified by Means-Ends Descriptions or Task Analyses ("What Actually Happen," Chapter III). In a Means-Ends Description we merely specify the Elements of the process (or of each of the Stages) which contribute to the outcome of the process (including desired changes, if any). In a Task Analysis we merely specify what sets of states of affairs would qualify as successful Outcomes (in effect, we specify different 'Versions' of a successful outcome).

Those who are familiar with the range of conceptual-notational devices for real world representation will recall that a Process Description is appropriate for representing actual processes both in the abstract and in various historical exemplifications. A process representation is not *generally* the device of choice, for example, if one wishes to represent machinery or "systems" of any kind. For those, a Configuration Description, based on the Basic Object Unit would be the appropriate resource, for in such configurations the structure takes priority over the processes it undergoes in that, in general, without the structure there would be no such processes. And again, conceptual schemata are not descriptions or statements about how things are. Rather, they must be used in making

constructions which can be used to give descriptions or make statements, but conceptual constructions can be used in many other ways.

IV. The Simulation Paradigm

This research paradigm is designed to enable us to (a) formulate our understanding of a phenomena (and/or our guesses about it) by generating a process representation of it and (b) test that understanding by predicting certain facts about the phenomenon on the basis of other facts. What makes these predictions possible is the structure of the process description including, importantly, Co-occurrence and Attributional Contingencies. Because of the predictive implications of the contingencies which are stated in "if-then" form, the obvious technical implementation for this research paradigm is computer simulation, hence the designation "Simulation Paradigm."

In the context of the BPU process representation, the Simulation Paradigm involves the following features.

- (a) We begin with the general notion of a process and ask, "What is the phenomenon?" The answer here is given by specifying it as a process, e.g., "providing (certain) mental health services to children," and distinguishing various possible exemplifications (Versions) as categorized by the investigator. Classes of exemplifications correspond to the "outcome variables" of the classic evaluation design.
- (b) We ask, "What makes a difference in how the process goes?" The answer is given in terms of (1) the parameters of the process or any of its elements, and (2) the values of those parameters. The first of these two is a set that corresponds to the interventions or other predictive (causal or noncausal) variables in the traditional outcome study. There is no substantive limitation on the type of parameter which may be in-

volved. Thus, aspects of institutions, programs, persons, and situations or relationships may all enter the picture as "what makes a difference."

(c) We ask, "Where does it make a difference?" In answer, we merely specify a locus in the process representation. Whatever makes a difference has to make a difference somewhere.

(d) Finally, we ask, "What difference does it make there?" The answer will have the general form, "Depending on whether the value of parameter Q is x or y, the exemplification (of, e.g., "providing (these) mental health services to children") will belong to classes A, B, C," There is no restriction on the kind of functional relation which may appear here. (The statement of these relations is an integral aspect of the Process Description; formally, it consists of specifying attributional and co-occurrence contingencies.) For example, it may be a simple linear function, but it may be a non-numerical decision table or a logical "either . . . or . . . and . . . unless . . . if" kind of function. The latter is of particular interest in connection with the problem of strongly interacting variables (it is the potential for non-numerical values which leads us to speak generally of "parameters" rather than the traditional "variables"). In this case, the outcome is expressed as an explicit joint function of the values of the several "interacting variables," and that function will commonly have "or," "unless," and "if" components. In some of the very simple special cases, the joint function will be identical to a representation within the familiar linear additive model or other models used in

multivariate analysis. (Any type of multivariate analysis can be accomplished within the Simulation Model.)

The simulation model has the following features of interest.

- (a) It permits (indeed, requires) explicit statements of functional relationships between predictor and outcome parameters, but "default" relationships may be entered in the absence of relevant information or hypotheses.
- (b) There is no limit to the form which these functional relationships may take or to the nature of the parameters involved (e.g. personal, institutional, physiological, etc.).
- (c) It permits a clean separation between the conceptual, psychological model of the phenomenon and a mathematical model of the statistics.
- (d) It permits the detailed examination of the effects of any subset of predictor parameters, since the remaining parameters may be held constant, e.g. by using group means for individuals, hence allowing no variation in that parameter, or by using default relationships. (Such analysis would normally call for computer simulation implementation.)
- (e) Because it is, in effect, a predictive test of the entire set of relationships simultaneously, the sample size required for testing does not increase exponentially as it does in 'purely empirical' multivariate designs. (A complete factorial design, far from being purely empirical, represents the a priori application of a mathematical model. Yet it is one of the favorite paradigms of practitioners who insist they are "merely finding out" about the phenomenon, or "letting the data speak for itself.")

(f) There are two characteristic features of the paradigm which are worth noting. The first is that the actual Process Description will in general have to be relatively accurate in order to be of practical value in the sense of generating predictions which are both accurate enough and differential enough. This does not imply, however, that it has to be complete in the sense of incorporating all the concrete details of the phenomenon; since the Process Description is a "top down" approach one only goes down to the level of detail which is needed (or which is available), and, as we have noted, default relationships can be introduced where information or hypotheses are lacking. Thus, this approach makes possible simulation studies which are not possible under the traditional conventions of "bottom up" construction which do require all the detail. For example, in program or organizational evaluation we can generally obtain the required degree of information by interviewing a number of knowledgeable people; in traditional Evaluation Research, simulation requires so much information that, essentially, it is impossible to do.

The second is that in simulation studies there is no built-in analytical procedure for establishing the 'best fit' to a set of sample data. (Note that what we ordinarily call "best fit" in academica is almost wholly the product of convention. It is the best fit, given (a) a conventional criterion, e.g. the least squares criterion and (b) a conventional a priori, non-psychological model, e.g. a linear-additive model.) Inspection methods and systematic variation of hypothesized relationships are available, however, and

post hoc revision with cross-validation is a standard procedure. These latter methods provide for a much richer analysis of a set of data than is possible within the limitations of standard experimental designs and the associated statistical analyses. Traditional techniques are special cases within the range of possibilities generated by the simulation paradigm, since the linear additive model or any other mathematical model can be written in the form of functional relationships which are part of the simulation paradigm. Correspondingly, traditional statistical and curve-fitting analyses will be special cases of the "systematic variation of hypothesized relationships" noted above.

Here, as in general, the thrust of the Descriptive formulation is not to deny the possible utility of well known statistical and mathematical procedures. Rather, it is to provide a sufficiently general and fundamental methodology so that we can suit procedure to purpose rather than being constrained to a blind, uncritical, or fanatic adherence to the current folkways of academica to the detriment of our understanding of our subject matter.

- (g) Those who are familiar with the customs in Descriptive Psychology will almost automatically think of the preceding presentation as a Paradigm Case Formulation. There are clearly all manner of variations and extensions to be had on the basis of the Simulation Paradigm presented above, and there is no need to belabor the obvious.

There is one extension which may be worth noting. In What Actually Happens, I noted that "systems" representations will be given

by particular cases of Configuration Descriptions, i.e. Process Descriptions with Object components or Object Descriptions with Process components. In the case where the object of our study is a child's first year in school, a Process representation is clearly the most direct form of representation; the relevance of relationships among the child, the teacher(s), the parents, the other children, et cetera can appear in the form of Contingencies and functional relationships. In contrast, if our interest was in the classroom or the school or the school system as a system, an Object description would be the most direct formulation and the relationships among child, teachers, peers, supervisory staff, parents, etc., would appear explicitly within the Object description. Even here, however, it is the processes which take place involving the Elements of the system which would be the primary vehicle for simulation studies. Thus, the logic of the simulation paradigm is easily extended to the study and representation of systems of various sorts.

V. Evaluation Research

Evaluation research and routine program evaluation occur within the context of some basic facts which impose strong limits on the manner and ease with which evaluation can be effectively accomplished and the uses to which a given evaluation can legitimately be put. Among these brute facts are the following.

- A. Improvement does not occur in "pure form." Just as an actual success is never merely a success but is also a particular accomplishment such as

- winning a race, so an actual improvement on the part of some particular individual is always also a more specific change in personal characteristics, behavior, relationships, achievements, etc. But it is the fact of improvement that we are interested in, not the particular changes per se.
- B. There is no specific change which is per se necessarily or intrinsically an improvement. The decision as to whether a given change will qualify as an improvement requires evaluative human judgment. This judgment will, in part, reflect the target individual's personal characteristics and his life setting, as well as the purposes or norms with respect to which the issue of improvement arises for the person making the judgment.
- C. When improvement is exhibited by different individuals, in general, it is exhibited in different specific ways, even when it is "the same" improvement. (The differences will correspond to the differences between different Versions of the same process.) Thus, in principle, there is no way to specify rigorously and in objective terms the specific changes which would qualify as improvement for an entire group of individuals. At its worst, such a stipulative attempt would approximate the irrelevance of flipping a coin to decide the question of improvement. At best, it might be good enough for some purposes for some people.
- D. In principle, there is no change of a specific sort which all observers would agree qualified as an improvement for a given individual. In general, a given change in an individual will affect different significant persons differently and they will, properly, make different judgments of "improvement." Judgments of this sort will, therefore, carry weight only among persons who are in agreement in this respect. The likelihood

of agreement may be increased in several ways:

1. By moving towards "lowest common denominator" phenomena: "If he's learned to dress himself (or "if he can now hold a job," "if she can now sit still in class") surely that's an improvement."
2. By specifying norms or frameworks: "From the point of view of reality contact, his being more aggressive at work is an improvement."
3. By restricting the implications of the judgment or uses to which it is to be put: "For purposes of deciding whether to terminate."
"From the standpoint of how to advise the parents," etc.
4. By stipulating that one party, e.g., the client, the parent, the school or agency representative makes the decision.

It is easy to assimilate the foregoing to traditional views and hence to misunderstand its import. Thus:

Wil: You're telling me that it's impossible to specify evaluation criteria in advance. But in that case all you could provide with your so-called evaluation research would be post hoc, accidental, arbitrary, and probably self-serving conclusions. That would kill evaluation research as a scientific enterprise and put it on a par with reading tea leaves.

Gil: Hold on! The sky isn't falling. I didn't say you couldn't specify evaluation criteria in advance. Indeed you can. What you can't in general do justifiably is specify criteria independently of context. That would be as foolish as trying to specify what things would actually be dangerous independently of context.

We manage the concept of "danger" quite well, thank you, even though our judgments of what is actually a danger to whom are context-bound, and the same for concepts like "good," "successful," "valuable," et cetera, which are involved in evaluation. Being context-bound is not a handicap: that's the kind of concept they are. If you think that specifying criteria in advance depends on specifying criteria universally, which in turn depends on specifying them in a context-free way, then you're in trouble. Your own rationale makes what you do illegitimate.

Wil: But if not on those grounds, then on what basis can one specify criteria in advance?

Gil: On the basis of being knowledgeable about the situation and about what is at issue, keeping in mind that different decisions in the matter are possible, because for different people or from different viewpoints there are different interests and different issues involved. Good bureaucrat that you are, you have a passion for uniformity and formality, and a distaste for simple reality. But passing regulations as to which things are to be called "dangerous" and which things are not will not keep you from getting killed off by one of your 'non-dangerous' things, nor will setting up arbitrary conventions about which criteria are the mark of success in organizations, programs, treatment, et cetera succeed in making silk purses out of sow's ears. As a bureaucrat, you can set up those conventions for your purposes, but as someone who has to cope with actual clients, actual program effects, et cetera I can't afford to fool myself

that I can define good results into existence just by calling them that.

If I don't discount my knowledgeability, I am quite willing to say in advance, "In this situation, for my purposes, X will count as a warning of possible failure and Z will count as a definite failure, et cetera." What would be plain stupid is (a) to say "In all situations, for every purpose, X will count as the mark of success, etc." or (b) to suppose that I need a universal of that sort in order to make or justify my judgment on a particular occasion. As a bureaucrat you're supposed to follow orders and not do anything for which you could be held personally responsible. As a citizen I have to make decisions and take actions for which I am personally responsible. You live a very sheltered life.

But also, if the situation calls for it, I am quite willing to say after the fact that I was mistaken in using X as the mark of success because, as it turned out, there were identifiable factors in the situation which made X unsuitable as a mark of success. Both that decision and the decision in advance to count X as a mark of success are subject to challenge and justification. I wouldn't try to evade the responsibility by quoting regulations and pleading that I was only following my orders.

E. In a real world setting the effects of a given influence, e.g., mental health treatment, are inevitably confounded with an unknown number of important events and other influences. Changes from before to after cannot simply be attributed to a given influence. The experimental devices of random assignment to groups or extensive control of individual cases are palliatives, not solutions, and they are almost never available in field settings, since therapeutic, administrative, political, or other considerations will properly take precedence.

It is in part because the effect of such extraneous influences is subject to re-evaluation that the choice of criteria is also subject to re-evaluation. The simulation paradigm provides an explicit framework for taking such influences into account insofar as we actually can, both before and after the primary data collection. By doing so it sensitizes us to this kind of possibility even when we do not represent the "influence" in question in our simulation. For example, I might designate "getting and holding a job" as a mark of improvement, and count a particular set of clients as not having improved because they do not meet this criterion. But if economic conditions have substantially worsened and unemployment is substantially higher now than previously, am I just being arbitrary in deciding to give that 'criterion' less weight now? Should I have foreseen this possibility in the original selection of criteria? (Recall the issue, above, of taking all possible precautions in advance.)

Perhaps it is easier now than previously to see the danger of talking in terms of "criteria" here. The term suggests both something

separate from the phenomenon and something foolproof, and the combination creates problems. (a) We do not normally think of a phenomenon as its own criterion, so that reference to a criterion suggests that we have to go by something other than the phenomenon of interest. But what is at issue is whether the clients have improved. If getting and holding a job is involved in having improved, it would make some sense to speak of that as a "criterion" but then we might better just speak of their having improved. If getting and holding a job were only empirically connected to having improved, it would be misleading to call it a criterion. In Descriptive Psychology we will generally refer directly to the phenomenon or to (mere) indicators rather than to "criteria," and the choice among them will codify the nature and extent of our willingness to change our minds. This is preferable to the illusion or pretence of certainty carried by "criterion."

- F. Evaluative phenomena, being historical rather than universal, will, in general, be different across time and place. A treatment procedure which is effective here now is likely not to be effective somewhere else (different clients, different problems, different milieu) or here ten years later (different staff, changing times leading to changing problems, etc.). Thus, the utility of a given evaluation is limited. In an organization, evaluation is likely to be required on a continuing or periodic basis.
- G. All of the foregoing were phrased in terms of the evaluation of "improvement." Corresponding statements could be made in respect to other evaluative concepts such as "adequate functioning," "psychopathology,"

"mental health," and "life problem." There is no set of behaviors, objective indicators, et cetera which is logically equivalent to any of these evaluative phenomena.

- H. There is no set of procedures which intrinsically or necessarily qualifies as performing an adequate program evaluation. The decision as to whether a given set of procedures does so qualify on a given occasion requires human evaluative judgment. The adequacy of a program evaluation procedure provides the same general problem of evaluation as does the degree of improvement of clients under treatment.

Consider the case of new programs for children and elderly persons initiated by an urban mental health center. The systematic evaluation described below is designed to be responsive to the general considerations noted above and to be of practical value for a variety of purposes, including decisions regarding changes in program procedures and decisions regarding allocation of resources to different aspects of the Child and Elderly programs.

A. Evaluation Structure

The overall evaluation has three major components, which are here designated as predictive, procedural, and retrospective.

1. The predictive component resembles the classic outcome of treatment design in which indicators of improvement are designated initially and are assessed before and after treatment.
2. The procedural component is related to traditional "process" studies of psychotherapy in that it involves a process description of the

"treatment" or "program" and an evaluation of its appropriateness independent of outcome.

3. The retrospective component is a closure-achieving procedure. It involves:
 - a. An evaluative review of the original assessment of the treatment planning;
 - b. An evaluation of improvement unrestricted by the predictively designated indicators of improvement; and
 - c. An evaluation of the extent to which improvement or lack of improvement could be attributed to factors other than the treatment program.

B. Evaluation Procedures

The following is a narrative outline of the procedures involved in the evaluation. The integration of treatment and evaluation is such that all but the last section (the retrospective components) is also essentially an outline of the treatment process.

1. Intake

In this phase, an intake worker obtains standard background information and a statement of the problem, and makes a routine assessment of personal characteristics, including behavioral tendencies and personal resources and deficiencies.

2. Treatment Planning

An ad hoc treatment team is formed for the client. The treatment team will include at least (a) the primary therapist or caregiver, (b) another clinically knowledgeable person, (c) the continuity of

care person, and (d) for designated cases, a member of the evaluation team. The treatment team reviews the assessment information and does the following.

- a. Requests further assessment, if needed
- b. Judges the degree of need for treatment ("severity")
- c. Decides on strategy and type of treatment
- d. Specifies the nature of the problem
- e. Specifies the particular ways the problem is manifest ("ad hoc indicators")
- f. Specifies prima facie indicators for improvement. It is primarily the "severity" rating and these indicators which appear in the predictive "before and after" analysis. Among the indices which are likely to be used are those below. Note that these will in general be different for each client.
 - (1) Specific achievements, e.g., expresses affection toward her son, establishes a friendship with someone, learns class lessons without disrupting class.
 - (2) Changes in personal characteristics, e.g., becomes more responsible, less passive-aggressive, more tolerant of other people's shortcomings, less anxious, etc. Along with the specification of such changes is a specification of the preferred way of establishing these changes, e.g., therapist judgment, self-report, standardized test, special interview.
 - (3) Judgments by two significant figures in the client's life in regard to either characteristics or changes. Among such possible significant figures are the client, the therapist, a

family member, an employer or teacher, a friend, a spouse or spouse-equivalent, an agency representative. Part of the specification here is what judgments on the part of those persons would confirm the evaluation of improvement. (In light of the introductory comments, it may be noted that sometimes the confirming indicator would be a negative judgment by a significant figure.)

(4) The length of time required for effective treatment.

- g. Specifies, if possible, a set of criteria or prima facie indicators for classifying treatment as successful and for terminating treatment accordingly.

3. Treatment Review

At least once during treatment (perhaps at an interval of one-third the time estimated for effective treatment) the treatment team reviews the treatment procedures. The review is based on the treatment plan and the therapist's progress notes together with any supplementary information which may be available from the evaluation team. At this time, the treatment team decides (a) the degree to which treatment is in accordance with the plan, (b) the degree to which the treatment now appears to be appropriate, and, if indicated, (c) how to proceed with a modified treatment. The evaluation, together with a parallel evaluation by the evaluation team after termination constitutes the procedure component of the evaluation.

4. Termination Evaluation

This evaluation is conducted by the treatment team and may be initiated by either the primary therapist or the continuity of care per-

son. The evaluation is based on the treatment planning and review data, summaries by the primary therapist and the continuity of care person, and assessment (of improvement indicators) data provided by the evaluation team member. The treatment team makes a new rating of need for treatment, reviews its previously formulated criteria for success, and recommends termination or continuation. In the latter case the procedure is repeated when termination is again in question.

5. Follow-up

At a suitable interval after termination, normally six, nine, or twelve months, the indicators of improvement, or a subset thereof, are again assessed. A comparison of this data with the termination assessment provides a descriptive characterization of the stability of improvement.

6. Retrospective Review

The retrospective review is conducted by an evaluation team which is formed ad hoc for each client. Only the continuity of care person will be common to the treatment team and evaluation team. The evaluation team will include at least (a) a member of the evaluation staff, (b) the continuity of care person, and (c) a mental health professional who would be competent to serve as the primary therapist for the client in question. The evaluation team will have access to all the information available for the given client. The evaluation team makes judgments in regard to the following.

- a. The appropriateness of the assessment procedures and of the diagnostic conclusion or problem formulation.
- b. The adequacy of the treatment plan.

- c. The validity of the improvement indicators.
- d. The degree of need for treatment at termination (independently of the treatment team decision and the predictive criteria of of success and indicators of improvement).
- e. The degree of improvement shown.
- f. The extent to which improvement or lack of improvement is attributable to treatment or other influences.
- g. A post hoc reformulation of problem formulation and treatment plan. ("If we'd known then what we know now, . . .") for reference and future practice.

C. Integration of data

Each of the before and after indicators of improvement will be obtained directly in quantitative form or will be transformed into quantitative form, so that the before and after differences may be evaluated as to their statistical significance by means of, e.g., t-tests. Since these are only indicators rather than genuine criteria, these results will be only suggestive rather than decisive (fortunately, since they are likely to be contradictory).

The pattern of results which would most clearly support an overall evaluation of "improved" or "successful" would be the following.

1. All the improvement indicators show a statistically significant positive change.
2. Severity ratings show a change from "needs treatment" (to some degree) to "doesn't need treatment."
3. Either all significant figures agree that improvement has taken place or else the pattern of disagreement among these judges is predicted and explained adequately.

4. The treatment plan was implemented and appropriate, as rated in the treatment review procedures.
5. Retrospective review specifies that
 - a. Assessment and treatment were appropriate.
 - b. The client improved and was not in need of treatment at termination.
 - c. The client's improvement is attributable primarily to treatment rather than to other influences.

The major likelihood, of course, is that the results will be more or less equivocal rather than conforming to the "ideal" case above. Thus, the final evaluation problem will be how to count different patterns of less than completely unequivocal results. No rigorous general principle or procedure for accomplishing this evaluation is possible. It is possible, however, for the evaluation committee to formulate ad hoc procedures for generating composite "degree of success" indices for particular purposes. (This is in accordance with the principle of limiting use of results as a way of increasing the likelihood of agreement.) It may also be possible to specify general categories of data which would have utility across different evaluations. Such categories would be useful for summarizing the results of several evaluations or for comparing them.

D. Modifications

The general structure and procedure described above are primarily designed for the evaluation of improvement of individual clients undergoing comprehensively planned treatment for a substantial period of time and under conditions ideal for evaluation. Certain modifications would be

called for under other conditions. Among the modifications of present interest are the following.

1. Conditions for evaluation are almost never ideal. Because of the amount of professional time involved in the evaluation procedure it almost certainly will not be feasible to carry out all of the procedures with every client. In that case it would be appropriate to select every nth client at random for the complete evaluation procedure. A ratio of one in five or one in ten would seem to be reasonable in this regard.
2. Some telescoping of the procedure would be required in cases where treatment was not continued for a substantial period of time (e.g., on site treatment or early termination).
3. The evaluation of the Child and Elderly programs has both qualitative and quantitative aspects. The evaluation design described above is primarily responsive to the qualitative aspect; if clients improve, then, qualitatively, the program may be judged to be a success.

To a large extent this evaluation is an exemplification of the Precaution Paradigm. The whole notion of evaluation research, as it has evolved, is that one needs to take precautions against (a) wrongly assuming that what one is doing is successful, (b) continuing to invest resources in activities which are ineffectual, (c) failing to take advantage of opportunities to improve performance, or (d) other ways of going wrong, or conversely, that one ought to proceed with some assurance that one is doing it (whatever "it" might be) correctly, completely, successfully, et cetera.

Thus, for example, the details of intake and treatment planning (1 and 2, above) reflect the following concerns, precautions, and reassurances.

- (a) The initial assessment provides some assurance that treatment planning can proceed in accordance with local norms. Planning treatment at all is a precaution against providing inappropriate treatment.
- (b) The composition of the review team provides some assurance that the relevant interests and points of view (including those of the evaluator) will be reflected in the decision making.
- (c) The procedures of the treatment team constitute several precautions against inappropriate treatment.
- (d) The procedures of the treatment team provide some assurance that data needed for evaluation will be available.
- (e) The ad hoc character of the team (constituted specifically for each client) is a precaution against the bureaucratic pressures and tendencies toward uniformity. Procedures (f) and (g) assure that data for the prospective study will be available and provides some assurance that the relevant observers (including the care-givers) will be sensitized to recognize improvement or lack of improvement and that, correspondingly, their eventual judgments in this respect are less likely to be capricious.

In the review procedure (3, above) Precaution and Simulation Paradigms are both reflected.

- (a) The introduction of a review and its timing in the treatment process constitute a precaution against (1) letting inappropriate or ineffective treatment continue longer than necessary and (2) failing to take advantage of opportunities to improve effectiveness.

- (b) Reviewing the process provides some assurance that the conceptualization of the way in which improvement could be expected to take place was sufficiently cogent and accurate to warrant the confidence placed on it.
- (c) The use of the Simulation Paradigm would be an explicit and systematic elaboration of the preceding point. The process in question would, say, be the process of therapeutic interaction. Various classes of Versions, including those corresponding to successful and unsuccessful outcomes could be distinguished, and therapeutic and non-therapeutic contingencies formulated, and so forth.

The termination evaluation procedure (4, above) provides some assurance that the various relevant interests and viewpoints will be reflected in the termination decision. The reference back to the previously formulated 'criteria' for termination is a precaution against capricious decision making (for example, it recalls for the team the *basis for* choosing those 'criteria'). Reference to the process review provides some assurance that the apparent improvement is consistent with the reasons for expecting it. The composition of the team is a precaution against self-serving evaluation or merely self-fulfilling treatment planning. ~~The~~ ^{ff} follow up (5, above) is a precaution against premature judgments of sufficient improvement.

The retrospective review (6, above) is a precaution against making or having made incorrect or less than optimal decisions with respect to the client or with respect to the evaluation. It also provides some assurance that experience with this client will contribute rationally to future practice. (Appeals to sampling theory and 'generalization')

for this purpose in this context would generally be merely bizarre.) It provides some assurance that all the relevant facts, and not merely those which were initially anticipated, are taken into account in arriving at conclusions which have a rational prescriptive value (in the Critic sense) for further action.

The evaluation research practitioner who is thoroughly steeped in dolce academica will v^ery likely protest that I have failed to distinguish 'carefully' between the program and the evaluation. That is not accidental. I have tried to illustrate the general principle that the difference is one of function rather than of procedure as such and that in the limiting case there may be complete overlap between the two. Our social practices did not evolve and are not exemplified in an ecological vacuum, and our participations in them have no a priori guarantee of success. Consequently, our participations in those practices do not, in general, consist of plunging blindly ahead (pure Actor functioning), so that a very special and categorically distinct set of activities is required to set us on the right path or keep us on it. Rather, precautions, assurances, and self-regulation are general prerequisites for normal human behavior. If prudential considerations often call for a greater degree of explicitness and systematization than is "built in," that could provide a *raison d'etre* for a distinctive discipline called "evaluation research." But the rationale for evaluation research lies in the logic of precaution-taking and assurance-seeking and self-regulation and not in the transcendental visions of sampling theory, 'experimental design,' et alia.

The research design described above is, at the given level of description, more rigorous than any actual evaluation research known to me and more rigorous, flexible, and resource-full than commonly accepted

examples and norms for "good practice" in evaluation research. For example, most evaluators would not consider the possibility of separate prima facie indicators for each client, either because that would violate a current interpretation of "objectivity" or because it is assumed not to be situationally feasible, et cetera. Too, most "good" evaluation designs do not include the retrospective component, probably because that would run counter to the predictive conventions of 'experimental design' or 'the scientific method.' And again, current evaluation technology allows for multiple outcome measures only insofar as they overtly agree; genuine and legitimate disagreement is so far a methodologically unmanageable concept. In contrast, the methodology of disagreement is a routine aspect of Descriptive Psychology.

But, to repeat, the point of the present report is not to contribute new techniques to a field of endeavor already burdened with a procedural orientation and a swollen grab-bag of procedure-schemas. Rather, it is to show how, at least within a Descriptive Psychology framework, it is possible to be methodologically grounded and procedurally justified. There is something beyond ad hoc and shamanistic appeals to authorities, "the literature," "the scientific method," or "customary good practice in evaluation research." To be sure, many current practitioners would say they see no value in such "philosophizing" unless it pays off in new procedures and statistically significant results. But it is characteristic of Descriptive Psychology to be concerned with such matters.

APPENDIX C

MATERIAL ON PARADIGM CASE FORMULATION
AND PARAMETRIC ANALYSIS

(Reproduced by permission of
Linguistic Research Institute)

Linguistic Research Institute

Report No. 22

CONCEPTUAL-NOTATIONAL DEVICES:

The PCF and Related Types

Peter G. Ossorio

1979

© Copyright 1979
ALL RIGHTS RESERVED

Linguistic Research Institute
Boulder, Colorado

When we acquire a concept there is an extended family of possible ways of acting on it. We may, for example, use it in recognizing instances or non-instances and treating them accordingly. We may also use it in hypothesizing, imagining, predicting, wondering, instructing, speculating, asking, wishing, demanding, denying, pondering, or asserting. And more. In general, the acquisition of a concept opens up behavioral possibilities. It adds to our behavior potential.

For example, the acquisition of the concept of a Person gives us the behavioral possibilities of a person. (Compare: the acquisition of the concept of chess gives us the behavioral possibilities of a chess player.)

In Descriptive Psychology the formulation of the Person Concept and of more detailed subject matter is accomplished by means of a number of notational devices. One reason for this strategy is that since concepts have no possible truth value, statements are never involved in an essential way in the conceptual formulations themselves, but at most appear in some accompanying commentaries or illustrations. As a result, the possibilities of straightforward discursive presentation are very limited, and so some alternative is required.

Because notational devices of verbal and other sorts are public and communicable, they play an essential part in the public and communicable character of concepts.

Because of this essential connection, and also for certain heuristic purposes, the notational devices used in Descriptive Psychology are generally designated as conceptual-notational devices. We also distinguish between particular conceptual-notational devices and the more general types of which they are exemplars. The latter are designated as conceptual-notational device types.

Our present concern is with four related device types. These are (1) the definition, (2) the paradigm case formulation, (3) the parametric analysis, and (4) the calculational system. Various exemplars of each of these are found in the literature of Descriptive Psychology. One reason for considering these as a group is that each has some relevance to the problem of introducing a subject matter without any essential reference to any other subject matter or dependence on any other subject matter. A second reason is that these four device types are systematically related to one another. The domain within which these relationships have a place will require further elucidation at a later time. I expect to have more to say about it in "The Behavior of Persons."

I. Definition

To define a term, "X," is, traditionally, to specify the necessary and sufficient condition(s), Y, for the correct use of the term. Certain additional qualifications are involved.

- (a) The satisfaction of condition(s) Y must be what makes the use of the term correct. (Having an angular sum of 180° is a necessary and sufficient condition for a plane polygon to be a triangle, but it is having three sides that makes it a triangle and is the defining characteristic. One might put it that its being a triangle is its having three sides, whereas its being a triangle is only logically equivalent to its having an angular sum of 180° .)
- (b) The term "X" is to be used to refer to the states of affairs which consist of the satisfaction of the necessary and sufficient conditions. (Thus, even if one could [and we can't] specify the necessary and sufficient condition[s] for the correct use of the term "Aha!," that would not be a definition, because "Aha!" is not used to refer to those conditions. Here, "correct use" paraphrases as "appropriate" rather than "true.")
- (c) The term "X" does not appear essentially in the specification of the condition(s), Y. If it does, the definition is "circular," and one would normally say that it wasn't really (didn't do the proper job of) a definition, but rather, only had the form of a definition. (The reference to "essentially" will legitimize recursive definitions.)

Reflection on this last qualification provides an interesting reminder. For any term, "X," for which a necessary and sufficient condition for applicability is relevant, there is a simple, direct, and rigorous way of specifying that condition, namely "X." For example, the necessary and sufficient condition for the correct application of the term "blue" to a thing is that the thing be blue; for the term "angry," that the individual be angry, and so forth.

Hence a traditional definition is not the way of specifying the necessary and sufficient conditions for the correct use of "X." Rather, it is a second way of doing so. Not too surprisingly, a second way is usually not available. (Just as one may argue that no two words are really synonymous, so one may argue that no real world phenomena are really definable except certain matters involving explicit conventions, i.e., those which are, in effect, created by a definition. It is instructive to try to formulate a really rigorous definition of something like a chair, a mountain, or a lemon, just as it is often instructive to try to give a set of directions for even a simple behavior such as tying one's shoes such that if only those directions are complied with, the behavior must be successful. Such exercises can do much to clarify the distinctive contributions of knowledge and competence.) Thus:

(a) Definitions are given for the sake of a listener who is not already clear enough about the way "X" is used or is to be used. They are not in principle necessary for identifying the necessary and sufficient conditions for the correct use of "X," nor are they necessary for picking out cases of X.

(b) The clarification attempted by recourse to a definition may not be successful. In that case some further definitions of the terms used in the first definition may help. But this procedure cannot be carried out ad infinitum. Definitions can be given in this recursive fashion but the success of a definition depends ultimately on the successful use of terms which do not need further clarification for the task at hand. Either the original term "X" or any preferred definition is ultimately an appeal to the competence of the listener, not a device for creating something out of nothing. Thus, any general 'requirement' that all terms be given (or even be capable of) definition is incoherent.

(c) Unless one already understands "X" there is no way to tell whether the specification of necessary and sufficient conditions given by "X," i.e., X, is the same as the specification given by a definition of "X," i.e., Y. Perhaps more often than not, there is a difference. In the case where we detect a merely appreciable difference we will generally say either that the definition is faulty or that a new word has been introduced in a misleading and possibly deceptive way, since it is spelled the same as "X." In the case where the difference is fundamental, we will generally say that the definition is a reductive one and that a new word has been introduced in an apparently deceptive way since it is spelled the same as "X" and clearly means something other than "X," yet no disclaimer is offered.

For example, to define "pleasure" as "the diffuse experience of the operation of reward mechanisms" would be of this latter sort.

So also would be the definition of "verbal behavior" as "behavior the reinforcement of which is mediated by another person" and, likewise the definition of "love" as "the reciprocal satisfaction of dependency needs." These definitions are genuinely reductive, and not merely fatuous, because they take us from the original concept to a different logical category or to a different conceptual system (and in the terms of "What Actually Happens," to a different "world," or domain of facts).

Ordinarily, a reductive definition is offered when its author believes that the specification given by the definition covers the same set or range of cases as "X." However, a definition requires more than coincidence in this respect for in that case "theory of" rather than "definition of" would be the appropriate characterization. Thus, the test of the 'definition' is to offer a hypothetical separation and see which way the decision goes. Of the author of our "pleasure" definition we would ask, "Suppose you had a case which by all existing standards was a case of pleasure ("If ever there was a case of pleasure, this is one.") but also, you had established to your satisfaction that it was not a case of diffuse experience of the operation of reward mechanisms. Would you then say, 'Then it can't be a case of pleasure, because it's not a case of experiencing the operation of reward mechanisms' or would you say 'Well, I guess pleasure isn't, after all, the experience of the operation of reward mechanisms'?" In the latter case we would conclude that the 'definition' was a bogus one because, as it turned out, it was not, even for its author, a specification of necessary and sufficient conditions for pleasure.

In the former case we would conclude that he had deceptively introduced a new word, and an associated concept, into the language and was pretending that the concept was that of pleasure. Conversely, the author of that genuinely reductive definition would also, it would seem, be committed to the customary follow-through of reductive definitions, namely the claim that "Properly speaking, there's no such thing as pleasure. What we (naively) call 'pleasure' is nothing but the diffuse experience of the operation of reward mechanisms. That's what there really is in the world."

The further importance of this reductive position is that if the definition is followed by investigation the latter will not be the investigation of pleasure (but instead, the investigation of the diffuse experience of et cetera), but it will be so presented to the scientific and general public.

Empirical data is heavily in favor of the prediction that a 'scientist' who offers such a definition of "pleasure" has access to some conventionally accepted technical means for studying the diffuse experience of the operation of reward mechanisms and none for studying pleasure. The definition will, therefore, offer the same advantage as looking for the lost coin under the lamp post because that's where the light is. Of course, I should not want to deny that we can study only what we can study, and only in the ways we have available. (Indeed, this is merely an instance of a general principle codified elsewhere as Maxim 5: "if a situation calls for a person to do something he can't do, he will do something he can do.") What is regrettable is that to date we should have been so unable to study our

professed subject matter and that our accounts of what it is we do as 'behavioral scientists' and what it is that we have found out should have been so obtuse or misleading.

If we hold to the requirement that our subject matter really be the behavior of persons, then we will be properly skeptical of the possibility of introducing that subject matter formally (explicitly and systematically) simply by means of either general definitions (e.g., of persons or behavior) or a collection of more particular ones (e.g., of "pleasure," "love," "verbal behavior," etc.), this in spite of the fact that our strongest proclivity in the matter, inculcated by academica is to begin with definition.

For, on the one hand, it would be almost impossible for an informative definition not to be a reductive one, which would lose us our subject matter at the first step (as it has happened in academica). And, on the other hand, we ought to have at least a nagging sense, if not a positive conviction, that if our hold on the subject matter were so tenuous that we had to anchor it with a second specification we should have no business claiming to study it. Conversely, one can imagine, for example, a 'behavioral scientist' announcing "I'm going to study the behavior of persons scientifically. Of course, I don't know what that is; I wouldn't know a case of behavior as such if I found one, and I can't tell a person from an organism, but I've got the experimental method and the hypothetico-deductive method and the theory of measurement on my side, and by cracky, there ain't anything that won't yield to those methods! Just give me twenty million dollars, six thousand subject hours, one

hundred years, and immunity from prosecution, and you'll see, I promise you!"

But of course, this isn't how to go about it, and the example is somewhat overdone. The "know nothing" approach has been viable in the "natural sciences" because the subject matters there are essentially completely invented ones, and so there were no antecedent reality constraints on what we might say, e.g., about electrons, ion exchanges, red blood corpuscles, et cetera. In contrast, when we study the behavior of persons, the course of their development, and the differences among them, there already is such a subject matter. It is what we need a better grasp of, and no amount of successful bait and switch tactics by our 'behavioral scientists' will meet that requirement. If we freely invent subject matters by means of definitions or theories of "behavior," "verbal behavior," "pleasure," et cetera we should not expect to have then a better grasp of what it was that we needed to understand better.

II. Paradigm Case Formulation (PCF)

A paradigm case formulation is possible in many cases where, in some relevant sense, a definition is not possible. For example a definition is not possible when there is no second set of necessary and sufficient conditions the satisfaction of which is what makes the use of the term correct. (Wittgenstein's example of games is a familiar heuristic in this regard.) A paradigm case formulation is accomplished in two major steps:

I. Introduce a Paradigm Case (of X)

II. Introduce one or more transformations of the Paradigm Case

As in the case of definition, the Paradigm Case must be either directly intelligible or finitely explainable in order for communication to be successful. The Paradigm Case will directly identify some portion of the cases which are to be picked out. Each transformation will pick out additional cases. Each transformation may be considered to be a constructional instruction or an indirect description:

"Change the Paradigm Case in this way (the transformation) and the result will still (also) be a case of X." Thus, if the PCF is successful, the Paradigm Case and the transformations will, collectively, pick out all and only those cases that one wants to pick out, even if there is no second thing that those cases have in common. Therefore, a PCF will accomplish the identification of a subject matter (a range of possible cases) no less effectively than would a definition.

A standard example of a paradigm case formulation is the following PCF for the concept of a family.

I. Paradigm Case: A husband and his wife living with their natural children, who are a seventeen-year-old son, a ten-year-old daughter, and a five-year-old daughter.

II. Transformations:

T1. Eliminate one parent.

T2. Change the number of children to N , $N > 0$.

T3. Change the sex distribution of the children to any distribution other than zero boys and zero girls

T4. Change the ages of the children to any values compatible with the ages of the parents.

T5. Add any number of additional parents.

T6. Add adopted and other legally defined sons and daughters.

T7. Eliminate the requirement of living together.

T8. Add zero children if husband and wife are living together.

T9. Eliminate the requirement that the parents have the legal status of "married."

Note that constructing a PCF has a good deal in common with constructing a definition. For example, constructing a definition often involves some careful decisions and judgment in regard to what cases to include and what not to include. Correspondingly, what someone approves as a "good" definition someone else will disapprove as misleading, defective, or wrong. One constructs a definition in

the course of participating in some social practice(s) which give it a point or a purpose; depending on what one is up to, one or another specific construction may be correct, successful, adequate, et cetera. And to give a definition of a term already in use, e.g., "behavior," "emotion," or "family" is to run the risk of violating the existing use and hence to run the risk of degrading the language and misleading or manipulating one's audience. All of these possibilities are the case for paradigm case formulations, though the danger is not as great.

There are also important differences between a definition and a paradigm case formulation. First, of course, we have noted that a PCF will do the job of identifying a subject matter in circumstances where a definition would fail because there is no second set of necessary and sufficient conditions for the use of the term in question.

Second, a definition and a PCF are differently "keyed" to the behavioral context in which alone they are intelligible. A definition is a definition of a term. The concept which is to be associated with the term is identified by the specification given in the definition. The term can thereafter be used to mark that concept or instances of that concept. In contrast, a paradigm case formulation is a formulation of the concept or the range of its instances; it is not a formulation of the term. Thus, in connection with "family," a definition would be a definition of the term "family," whereas the PCF given above is a PCF of the phenomenon of "the family" or of the concept of a family.

Consequently, and thirdly, whereas it is almost impossible for a definition not to be reductive, if only in a weak sense and not in

the strong sense discussed above, it is essentially impossible for a paradigm case formulation to be reductive. A definition of "X," unless it is circular and hence defective, must be given in terms of something else, "Y" and "Z." Almost inevitably, one loses the X in favor of the Y and Z, because that is what X "really is," and, after all, if one gives a definition that implies that "X" was not a good enough way of saying what X really is. (I am speaking of definition primarily in scientific discourse, not in mathematics, logic, et cetera, where issues of reductiveness are not as salient.) If one is not prepared to sacrifice the X in this way, one normally does not define "X" but merely uses it to refer to X. In contrast, since a PCF begins with a genuine case of X and each transformation leaves us with something that is still (also) a case of X, we are in little danger of replacing X's with Y's and Z's.

Fourth, what is involved in a paradigm case formulation is not merely which cases are picked out, but also how that is done. In most PCF's the transformations reflect differences among cases, and these differences are at face value relevant and potentially important for the task at hand or in other contexts. By virtue of its structure, the PCF will often help to make clear why it makes sense to consider as one type of phenomenon a set of cases which are importantly different from one another and have no second and separately specifiable set of criterial attributes in common by reference to which we could codify the homogeneity of the set.

Thus, a paradigm case formulation will, paradigmatically, have some illuminating and explanatory power that makes it no less akin to

a theory or a conceptual analysis than to the usual sort of stipulative definition. Indeed, in the family example, a consideration of the ramifications of being legally married or legally adopted would take us very naturally into the whole area of kinship and kinship theory and social systems, so that these latter could be considered simply as extensions and elaborations of the PCF for the family.

Correspondingly, the PCF lends itself readily to certain research designs. Since each transformation both picks out a group of cases and reflects a possible important difference among cases, it would be natural and sensible on the whole to study each such group separately in order to decide whether the empirical regularities which were characteristic of one such group were equally characteristic of each of the others and of "the family" in general. [This is the PCF stratified sampling design.]

One of the reasons why this research design has a methodological interest is the issue of how usable one's results are by someone else in some other circumstances. And one reason why this is as important an issue as it is is that people will disagree. The issue of reductive definitions is an extreme case here, and there are less extreme cases as well. Looking back at the paradigm case formulation for the family, it seems pretty clear that few consumers would disagree with the Paradigm Case and the first four transformations but that T5 through T9 would each be likely to generate appreciable disagreement or dissatisfaction. For a consumer of research who, say, objected to T5, having the results stratified along these lines would permit him for his own purposes to discard the empirical

results for that group and make use of the remaining as findings about "the family" while still being able to understand and communicate effectively with someone who didn't reject T5. Neither a definition nor simple random sampling or 'representative' sampling would permit this happy result.

There is more to be said in connection with the fact that a paradigm case formulation involves issues of how cases are picked and not merely which cases are picked. Perhaps it is clear from the example of the family that in general, in a PCF any of the cases of X could conceivably be used as the Paradigm Case and, depending on that choice, the set of transformations would be different. In that sense, the choice of a Paradigm Case is conceptually arbitrary. It need not be methodologically arbitrary, however. In general it will make a difference, and sometimes a crucial difference which one chooses as the Paradigm Case, and there are some reasonable rules of thumb which provide a basis for the advantageous choice of a Paradigm Case.

The first rule of thumb is to choose the most complex case as the Paradigm Case. The reason for this choice is that then usually the transformations are simplified, becoming merely a series of deletions or something close to that. When that is not the case, there is no great advantage in this choice. In contrast, starting with the simplest case is generally a poor choice for a Paradigm Case because then usually the more complex cases cannot be generated as simple transformations but instead will require substantive additions which present the same decision problems as in the original choice of

Paradigm Case. This is particularly so when the simplest case is also a derivative, or parasitical, case (see below). There is an obvious moral to be drawn here in regard to the classic theoretical-experimental strategy of studying artificially simplified cases in the 'laboratory' in order to "get at the fundamentals of the phenomenon."

The second rule of thumb is to start with an indubitable case. This was the operative rule in the family PCF. "If ever there was a case of a family, that's one." The virtue of such a choice, when it has a virtue, is that it makes disagreement more easily managed. I mentioned that a consumer who disagreed with T5 in the family PCF could for his own purposes throw out the results for T5 and use the remainder if the research were stratified in accordance with the Paradigm Case and the transformations. It would generally be more inconvenient and it would be difficult or impossible to make use of partial results if the consumer objected even to the Paradigm Case, for then there would be a good chance that none of the groups defined by the transformations would be cleanly acceptable either.

The third rule of thumb is to start with a case which is in some relevant sense primary or archetypal. (The rule is applicable only if there is such a case, of course.) The relevant consideration is that one wants to give formal recognition to the fact that the other cases are cases because of their relation to the primary case. One thinks, for example, of legal precedents. This case is a case of violation of privacy because of the way it is related to the case of *Dukes vs. Wisconsin*, but not vice versa. Or again, one thinks of

the case of perceiving objects and hallucinating or imagining them or the case of genuine coins and counterfeit coins. A defective, partial, illusory, counterfeit, or imaginary something is that because of the way it resembles the real thing and not vice versa. In connection with the PCF for the concept of a family, the Paradigm Case is primary relative to T7 (eliminate the requirement of living together). Given the paradigm case we can (especially with some additional restrictions) accept T7 as an exception or variation because of the way it resembles the Paradigm Case. In contrast, if all cases of families were (conceptually) cases of people who were not living together, that would not be our concept of a family. (Note that a definition of a family which was broad enough to allow for T7 at all would automatically allow the possibility that all cases of families were cases of people living apart.) Yet T7 is what makes intelligible our notion of a family with a daughter in college or a son in the army and our notion of an extended family (kinship system).

Considered as a formal device, i.e., a device type, rather than as a substantive construction, the paradigm case formulation has a reflexive use as well as a recursive logic. That is, not only is it the case that some element of a PCF can be given by means of a PCF, but also the very notion of a paradigm case formulation as given above can be handled in PCF fashion. Let us construct an example. I introduced the paradigm case formulation by specifying a two-step procedure, namely, (a) First, introduce a Paradigm Case of X, then (b) Introduce some number of transformations of the Paradigm

the conceptualization of the phenomenon involves T1 (an archetypal variety of masculine-feminine relationships) and T2 (actual relationships generated by non-actual archetypal ones) and the rationale for the use of numerical scales involves T1 and T3 as an alternative to the traditional theory of measurement. (This is an example of both a recursive paradigm case design and the non-mathematical use of mathematical procedures.)

(b) T4 permits us to assimilate "family resemblance" structures, which are just a little looser than the original PCF in that one generates the family by moving from case to case rather than always returning to the same case. This device type is available when there is no guarantee that all the members of the set can be generated from a single Paradigm Case.

(c) T2, T3, and T4 extend the scope of the PCF to cover generative "Rewrite" systems such as those found in much current work in syntax and semantics. An arbitrary rewrite rule, "P may be rewritten as Q" will in general not correspond to any independently specifiable transformation. However, by anchoring in this way on P and Q we may thereby define a new transformation ad hoc, i.e., the "P-Q transformation." Thus, the "rewrite rule" is the functional equivalent (T3) of a transformation. (It is also the functional equivalent of an Operation in a calculational system together with the eligibility constraint that the Operation can only be performed on the Element P—see below.) In such systems T4 is what would permit derivations (rewriting the results of a rewrite) and T2 is what would permit

sentences, or surface structures, to be derived from non-sentences, or "deep structures."

Surprisingly, perhaps, the concept of a paradigm case formulation has some appreciable stylistic significance in the presentation of the Person Concept. Very often it is important to stick to the theme or task at hand without getting bogged down in details, side-tracks, qualifications, et cetera. And yet, saying something in a relatively simple and direct way is likely to be ineffective if there clearly are qualifications, amplifications, exceptions, et cetera which would need to be taken into account somehow if one wanted to be both as complete and as careful as possible. Frequently one can (and frequently, I do) try to minimize that sort of problem by using "paradigmatic" or "paradigmatically" as a standard marker to indicate that what is simply presented is to be understood as the (or a) Paradigm Case (frequently, of the primary, or archetypal, sort) in a paradigm case formulation and that the loose ends are presumed or known to be manageable via transformations and their existence is not crucial for the task at hand (or that they will be or have been dealt with explicitly elsewhere). For example, I commented above that "a paradigm case formulation will, paradigmatically, have some illuminating or explanatory power that makes it . . . akin to a theory or a conceptual analysis." Of course, one can't guarantee that a given PCF will be illuminating; triviality and awkwardness are just as possible here as with definition. But on the other hand, merely saying that a PCF may be illuminating, but then again, it may not would be like merely saying that the members of a family may live

together, but then again, they may not. Were the both possibilities on a completely even footing a PCF would be importantly different from what I have presented and a 'family' would be essentially different from a family.

I might mention, too, that the present concept of a Paradigm Case Formulation has no interesting connections to the body of literature in philosophy, dating mainly around 1955-65, in which the terms "paradigm case" and "paradigm case argument" occur. There, the term "paradigm case" is used to refer to "the very kind of case from which one could (or from which we normally do) acquire the concept of X" (where "X" might be "red," "angry," etc.). In the present terms, such a case would seem to have some aspects of the "indisputable" and "archetypal" cases indicated in the rules of thumb for choosing a Paradigm Case. A paradigm case argument, roughly, is an attempt to establish the indisputability of calling an exemplar of the paradigm case a case of "X" by reference to either (a) its indisputability in the course of acquiring the concept or (b) its presumed archetypal character by virtue of which it could serve as an indisputable case in the acquisition of the concept. Fortunately, the concept of Paradigm Case Formulation presented here does not depend either on that use of "paradigm case" or on the success, failure, or meaningfulness of that argument. In general in those rare cases where a reference to "the very kind of case from which one could acquire the concept" is intended, that will be indicated explicitly.

Returning to the PCF: Having constructed a Paradigm Case Formulation, one could wave one's hand over it and transform it into the verbal semblance of a definition. For example, one could say, "By 'a family' I mean one or more parents living with one or more of their children." Such a 'definition' would need explanation, qualification, commentary, elaboration, et cetera. ("I didn't mean that no children was a borderline case." "Of course, a family with an only child away at college is still a family—by 'living together' I mean" "Of course, by 'parents' I mean either natural or adoptive parents, and by 'children' I mean either natural or adopted children, and . . . and . . . and") These emendations tend collectively toward a scattered and cluttered unwieldiness so that even with such elaborations the augmented 'definition' would lack the crisp, recursive logical structure which, as noted above, both illuminates the phenomenon and facilitates the stratificational research strategy and is essential for the archetypal and reflexive-recursive use. In short, for a definition to have the advantages of the corresponding paradigm case formulation it would have to amount to a PCF in form as well as content, and so its being a definition at all would be immaterial. Indeed, because a definition has the (paradigmatic) verbal form of specifying a single set of necessary and sufficient conditions, when it amounts to a PCF it will have to be considered a "loose" or "defective" definition (not a paradigm case definition or an archetypal one), since no amount of definitional syntax will create a second defining criterion where there is none. So much for scholarly niggling about whether we aren't "really" or

"after all" talking about a definition. Qualifiers such as "really" and "after all" are characteristically ways of expressing transformations of implicit or informal Paradigm Cases.

In sum, the concept of a paradigm case formulation is a distinctive resource for both thought and action. It has a perspicuous and deceptively simple logical structure. Partly because of this, it is, paradigmatically, a remarkably effective device not merely for introducing or locating subject matter but also for generating more elaborate constructions which are suitable for the conceptual and empirical mapping of domains of fact. It has a special value in those logically complex behavioral domains wherein the structure and regularities are archetypically psychological rather than merely mathematical, physiological, or anything else.

III. Parametric Analysis

Like a definition and a paradigm case formulation, a parametric analysis is a device for notationally identifying a conceptual domain, or range of cases. Like a paradigm case formulation and unlike a definition, it also provides a way of dealing differentially with different cases within that range or domain. Like a paradigm case formulation, and unlike a definition, a parametric analysis deals with a domain, not a term or locution.

To give a parametric analysis of a given domain of cases is to specify the ways in which one of those cases could be the same as another of those cases as such or different from it. (Such a specification will also permit us to specify the ways in which one

kind of case could be the same as another kind of case within the domain or different from it.)

The ways in which two such cases as such could be different from each other are exactly the same as the ways in which they could be the same as each other except for those ways in which they are necessarily all alike, i.e., whatever makes them cases within the domain at all. Each such way is a parameter of each case and of the conceptualized "general case." Because a parameter represents a set of possibilities, a parameter has formally associated with it a set of "values" corresponding to that set of possibilities. One specifies (more or less completely) which of these possibilities is the case in a given instance by specifying (more or less completely) which value the parameter has in that instance. Thus, paradigmatically, one picks out cases more or less uniquely within the domain by specifying, more or less uniquely, values for each parameter.

This general feature of the parametric analysis is directly reflected in the formula notation used here to represent a parametric analysis. An example of such a formula is the formula for the "general case" of behavior which is presented elsewhere:

$$\langle B \rangle = \langle I, W, K, KH, P, A, PC, S \rangle$$

In such a formula the left-hand term represents the domain, the right-hand term represents the parametric analysis, and the terms within the brackets represent the parameters of the phenomenon. Thus, as with the transformations in the paradigm case formulation, such a formula can be read as a reference to a general procedure: "To

specify something about one of these (B's) you have to specify something about some of these (the parameters)." One can also take it in a simply factual way: "When something of this sort (B) is the case, something of each of these sorts (the parameters) is the case." Or, in more 'dynamic' form: "When something of this sort (B) occurs, something of each of these sorts is the case." Or, in the idiom of sets, "The set of B's is the set of octuples I, W, K,"

A more familiar example of a parametric analysis is the "color pyramid," commonly found in introductory textbooks in psychology. The domain is the domain, C, of visible colors. In the color pyramid, colors are arranged in a three-dimensional order as shown in Figure 1.

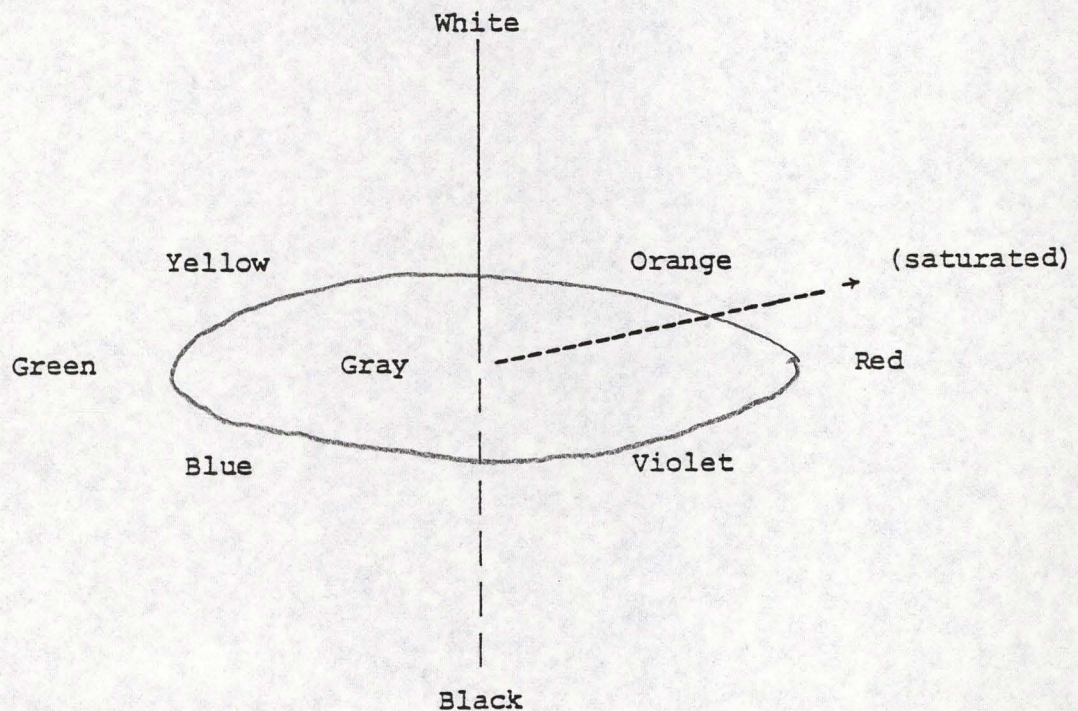


Figure 1. A Parametric Analysis of Color

There are three distinct orderings (parameters) in Figure 1. The first is the ordering from light colors to dark ones (White to Black). The second is the ordering from gray colors to brilliant, or intense, colors, as illustrated by the dotted arrow. (The actual ordering here is the perpendicular distance from the white-black axis.) The third is the familiar "rainbow color" ordering (circularly: red, orange, yellow, green, blue, violet). Ignoring some recently evolved niceties, these three orderings represent the ways in which one color can be the same as another color or different from another color. These three parameters are commonly designated as Brightness, Saturation (intensity, brilliance), and Hue. Thus, in formula form:

$$\langle C \rangle = \langle B, H, S \rangle$$

Because the concept of a parametric analysis is a purely formal (content-free) one there is no general restriction on, or prescription for, the kind of values that a parameter in such an analysis can have. For example, some parameters have numbers as their values; some have letters; some have facts; others have concepts; and so on. The only restriction is that all the values of a given parameter are of the same kind. (If it appeared that we had more than one kind of value for a given parameter, we would simply conclude that we were in fact dealing with more than one parameter or that we had conceived of the parameter too narrowly, etc.)

The nature of the domain is given in principle by what all the cases have in common. (Note that this no longer should suggest

that we ought to be able to define the domain.) Frequently, however, the distinctiveness of the domain is reflected in the distinctiveness of the parameters of the domain. Often enough, but less commonly, the distinctiveness of the domain is reflected in the distinctiveness of the values of the parameters of the domain. It is because the values of the parameters are in general the least distinctive aspect of the domain and because so much scientific effort is explicitly an effort to 'quantify' that we encounter so many cases where the parametric values are numbers.

The domain of colors is typical of a wide range of domains in that more than one parametric analysis is clearly possible. (For no domain is it demonstrably impossible that more than one parametric analysis could be given.) One could do it as in Figure 1, but one could do it by introducing any three-dimensional coordinate system, e.g., three orthogonal reference axes with an arbitrary base vector. Even if we consider only analyses in terms of brightness, hue, and saturation, there are still at least two ways of designating values for these parameters. The first is to provide categorical values, using qualifying adjectives (e.g., white, light gray, medium gray, dark gray, black), and this is typically done in ordinary discourse. (Grammatically, we normally use nouns for hues and adjectives for brightness and saturation if we talk about colors ["a light, toned-down green"], but we use adjectives for all three in talking about colored objects ["a light, gray-green chair"]; the latter reflects the fact that color is a parameter of material objects.) The second way is to specify the values of brightness, hue, and saturation with

numbers. This is frequently done in technical contexts, e.g., in manufacturing or in perceptual experiments. (Anyone who is bursting to ask, "But aren't you really just contrasting ordinal and ratio scales of measurement here?" is allowed to get it off his chest by writing it on the blackboard one hundred times.)

The two parametric analyses (categorical vs. numerical values) have different advantages and limitations. The advantage of the second, in principle, is that it opens up possibilities for numerical calculation (e.g., in predicting effects of color mixtures); it may or may not allow finer discrimination. The advantage of the first is that it is much more effective and efficient in picking out ranges of color which are in fact of interest and which would have to be represented in the three-dimensional space as irregular volumes. As Wittgenstein has reminded us, "Stand roughly there" is not a less precise instruction than "Stand exactly thirty inches north of that mark." Likewise, "a light, reddish brown" is not a less precise designation of a range of cases than "2.5-3.0; 3.3-3.7; 1.7-2.5." Indeed, picking out exactly that range of cases with a mathematical formula for an irregular volume in three dimensions is impossible, though we may jury-rig more or less adequate approximations. Note here the same kind of issue as with definition. Just as "pleasure" is not just a layman's naive way of referring to the diffuse experience of the operation of reward mechanisms, so "a light, reddish brown" is not a layman's naive way of talking about an irregular three-dimensional solid; rather, one might better say that the latter is the technician's naive way of talking about a light reddish brown.

The same considerations hold for "puce" or "the color of irises" and are even more to the point there.

Like the paradigm case formulation, the parametric analysis is essentially a non-reductive procedure. The specification of how one color may resemble another color or be different from it is, per se, not a reduction of color to something else. (But a parametric analysis may in turn be subjected to a reductive formulation, as when we say, "Hue is nothing but the wavelength of the electromagnetic excitation bouncing into your eye when you see the color.")

Like both a definition and a paradigm case formulation, the parametric analysis has a recursive grammar. The values of one of the parameters in a given parametric analysis may in turn be generated by a second parametric analysis. Indeed, our color example was just such a case. We noted that that color is one of the parameters of visible material objects, and it is the values of that parameter which are given by the parametric analysis of Figure 1.

A parametric analysis is more than just a substitute for a definition. It is also a preparation for dealing with the subject matter in a systematic way. In particular, patterns of various kinds or theoretical or empirical (real or hypothetical) regularities are potentially expressible as contingencies or other formal regularities or functional relationships connecting parametric values. For example, some important phenomena involving color can be represented as simply stated regularities in which the parametric values of a color mixture are expressed as a function of the parametric values and proportions of the component colors.

Such possibilities reflect the conceptual constraint that the only way a phenomenon or a historical particular can change is a change in the values of its parameters. Thus, it is unlikely that a sow's ear would change into a silk purse, but since they are both material objects, they have the same parameters, and so it could happen. In contrast, a mechanism could not become a motive, nor could an internalized parent or a history of reinforcement become a conscience or the ability to do arithmetic, and a flowerpot could not become the number 17.

Since a domain is at least as distinctive as its parameters, the latter are lower-limit determiners of what type ("logical category," "ontological category," "metaphysical category," etc.) of phenomenon (e.g., what type of object) the domain consists of. This is why, as indicated in "What Actually Happens," the specification of a set of "ultimate objects" also determines a self-contained "world" (a domain of possible facts), for the only possible happenings in that world will be changes in the values of the parameters of those objects and their aggregates as such, and the only possible states of affairs in that world will be the states of affairs identified by specifying the values of those parameters. In that world none of those changes or states of affairs will be the same thing as the existence of an object or process (etc.) of a different type. Only in the world of Persons is anything remotely resembling such an identity possible. Ironically, the major efforts of scientific and philosophical reductionists amount to an attempt to make the known world of Persons nothing but a world of non-persons, this under

curiously inverted theory of "naturalism." (See Sections V and VI of "What Actually Happens" for some elaboration.)

Like a paradigm case formulation, a parametric analysis is convertible into a "loose" definition. For example, "A color is anything that has some brightness, some hue, and some saturation."

But also, a parametric analysis and a paradigm case formulation will be formally inter-convertible. Thus, one could construct a PCF for color as follows:

I. Paradigm Case: light grayish-green

II. Transformations:

A. Change its brightness

B. Change its hue

C. Change its saturation

And the paradigm case formulation of the family could be converted to a parametric analysis along the following lines:

$$F = \langle P, NS, ND, LP, LC, A, AS, AD \rangle$$

where

F = family

P = Parents: M, F, M + F, M + F + O

NS = Sons: $N \geq 0$

ND = Daughters: $N \geq 0$

LP = Legal status of parents: M, C, U

LP = Legal status of children: N, A, O

A = Living arrangements: T, O

AS = Age of Sons: etc.

AD = Age of Daughters: etc.

In this analysis the values of the P parameter are (1) Mother, (2) Father, (3) Mother and Father, (4) Mother, Father, Other. The values of A are (1) Living together, (2) Other. And so on. Indeed, the 'definition' derived from the paradigm case formulation of the family ("some number of parents living with some number of their children") also corresponds to a crude parametric analysis in which the parameters P, NS, ND, and A are indicated but the nature of their possible values is left implicit (hence, in part, the need for commentary, elaboration, explanation, further definition, etc.).

Practically and aesthetically, paradigm case formulations and parametric analyses are not generally interchangeable, for all that there is a formal convertibility. A paradigm case formulation is, paradigmatically, suited to a situation where a certain general structure (including interrelationships among elements) is of interest. In contrast, a parametric analysis is more suited to a situation in which the range of variations among cases is of primary interest or in which a variety of patterns are to be mapped into the same domain. (This is not altogether dissimilar to the considerations affecting the choice of cluster analysis vs. factor analysis in multi-dimensional data-summarizing. Indeed, those who are familiar with these techniques may recognize them as hot-house varieties of paradigm case formulation.)

IV. Calculational Systems

There are some number of ways of delineating what a calculational system is. The following "Element-Operation-Product" model is the one used in explaining the system of reality concepts in "What Actually Happens."

One constructs a calculational system by specifying explicitly a finite set of Elements and a finite set of Operations. An Operation is something to perform on an Element or set of Elements, and an Element is something upon which to perform operations. If there are constraints on which Operations may be performed on which Elements, these constraints are also specified. When an Operation is performed on an element, the result is a Product. Whatever is a Product is also an Element, hence something upon which an Operation can be performed. Paradigmatically, the construction of a calculational system will include a distinctive notation for an Element as such (the canonical form) and for a Product as such (i.e. for a product as an Element-Operation combination), since calculation consists of systematically eliminating the latter by substituting the former. (The same kind of duality is found in the "Name" and "Description" structure of the descriptive formats presented in "What Actually Happens" for use in directly representing real world phenomena or for constructing more elaborate structures which can also be used for representing real world phenomena.)

A heuristic example of a calculational system (though mathematicians or logicians are likely to be unhappy with it) is given in Figure 2.

Element	Operation	Product
0 (original Element)	Add one (only Operation)	
0	+1	0 + 1 (1)
1	+1	1 + 1 (2)
2	+1	2 + 1 (3)
3		etc.
etc.		

Figure 2. A Computational System and Its Operation

In Figure 2, the Element "zero" and the Operation "Add one" are explicitly specified. The first Product is (zero)(Add one), which is the same as $0 + 1$, which is the same as the Element 1. When we perform the Operation "Add one" to this Element, we have (One)(Add one), which is $1 + 1$, which is 2. And so on.

Another heuristic example is the game of chess. For our initial Element we specify the initial board position of the pieces. For Operations we specify the rules governing the movement of each piece, specifying that that Operation can be performed only on that piece. (Note the resemblance here to the "rewrite rule.") Specifying that the Operations are performed on pieces is a convenience in form. In point of fact, the rules for moving pieces are rules for generating one board position from another by changing it in a particular way (moving a piece).

Computational systems have certain characteristic or distinctive features. Not every computational system generates an infinite set

of Products. (The "Add one" system above does this; the chess rules do not generate an infinite set of board positions, though they do generate infinite sequences of changes in board positions.) However, only a calculational system can generate an infinite set. Thus, our understanding and representation of limitless sets of numbers, sentences, descriptions, forms of behavior description, real world configurations, et cetera reflect the mastery of calculational systems. Note that our understanding and mastery, such as it is, does not come from generating an infinite set of products and inspecting the set empirically, but rather from our being competent to generate any one of the Products, not every one.

I commented above that the paradigm case formulation, paradigmatically, carries some element of explanation or illumination because it clarifies the relationships among a set of otherwise disparate cases having nothing else in common than being cases of X. To this we can now add that the understanding provided by a calculational system is an extension of the understanding provided by a paradigm case formulation, and this is so because a calculational system is an extension of a paradigm case formulation.

The initial Elements in the calculational system correspond to the Paradigm Case(s) in the paradigm case formulation. The Operation(s) in the calculational system correspond to the transformation, in the paradigm case formulation. The difference is that in the calculational system we can operate in principle on the results (Product) of an Operation (transformation). To get from a PCF to a calculational system, it is not enough to allow transformation on the results of

transformations ad hoc—it must be in principle. For example in the PCF for the family, we allowed the deletion of one parent. We could, if we were so minded, take the result of that transformation, i.e., a one-parent family, and allow once more the deletion of one parent. However, we could not allow such a deletion indefinitely or, indeed, even one more time. Since there are finite number of parents there are a finite number of deletions of parents that could be allowed (and note that because of that the successive deletions could be replaced by a single transformation, i.e., "delete any number up to $N-1$ of the N parents").

To repeat, our grasp of an infinite set of products comes not from enumeration and not from empirical inspection or inductive encounter, but rather from having the concept of the totality of things having a certain relationship (corresponding to transformation or Operation) to each other and to some paradigmatic cases of which we have some direct acquaintance. In "What Actually Happens" we noted that this kind of understanding is what makes both feasible and intelligible a behavioral science which deals with all behavior, actual and possible, and that from this consideration there follow some constraints on what would qualify as behavioral theorizing.

Returning to our comparisons, a calculational system not only can be derived from a PCF, but also, it can, with some possible awkwardness, be transformed into a parametric analysis. For example, consider a two-parameter analysis with parameters E (Element) and O (Operation) such that specifying parametric values (of the domain of Products) is the functional equivalent of specifying which

Operations can be performed on which Elements. The problem, in the parametric analysis, of specifying which sets of parametric values (which Products) were the values for a given Element would correspond to the problem, in the calculational system of specifying which calculations (sequences of Operations on Elements) would generate the same Product.

Finally, a calculational system, like both a parametric analysis and a paradigm case formulation, can be replaced by a loose (very loose) definition, e.g., "A chess position is any position generated by the 'chess rule' system."

Like both the paradigm case formulation and the parametric analysis, the calculational system as such is essentially non-reductive. This is because the Products are related to the original Elements via the transformations (Operations) and hence will, in general, belong to the same logical category. (Conversely, it is because (a) things of the kind defined by the original Element are in principle transformable, and (b) the result of transformation is again a thing of that kind that we can say that the transformations are repeatable indefinitely.)

APPENDIX D
THE LOGIC OF EVALUATION

Evaluation, or appraisal, is generally contrasted to description, but the nature of the distinction, the difference between the two, and the general character of each have all been formulated in a variety of ways. Most of these have been either unilluminating, clearly inappropriate, or positively confusing. Because of this, some conceptual discussion of evaluation as such is necessary in order to provide a clear basis for the technical and methodological formulations dealing specifically with the evaluation of information systems. The conceptual analysis comes from work in Descriptive Psychology (e.g., Ossorio, 1976).

One way to approach this pair of terms is by way of Piaget's (1952) well known distinction between accommodation and assimilation as fundamental aspects of behavior. In accommodation the organism does what it does because of the nature of the real things or circumstances it is dealing with. In assimilation the organism fits an external object, process, or event into some role within its activities and it is in terms of that role that it has reality for him. Inevitably, all actual behavior involves both accommodation and assimilation to some degree, but much behavior is characterized by the predominance of one or the other.

For example, in moving my hand toward the place on the table where the eraser is and in grasping it between my thumb and forefinger, I am primarily accommodating to the fact that it is where it is and that it is the size and shape that it is. Conversely, in seeing it as an "eraser" (which is a functional description already) rather than, say, "a piece of rubber" or "a small solid rectangle," and in using it to remove a word from the page I am writing, I am assimilating it to my activity of writing and editing.

One could say that I describe the object as "a piece of rubber" and evaluate it as "something I can erase with," and if I want something erased I will treat it accordingly, namely, use it to erase with.

With these preliminaries we may move to the more systematic definitions which follow.

1. An appraisal, or evaluation, is a description of something (or, correspondingly, a way of understanding something) which implies a certain motivation (but not necessarily overt behavior) on the part of the describer.
2. A mere description is a description of something (or a way of understanding something) which does not imply motivation on the part of the describer.
3. A descriptive appraisal implies the appropriateness of a certain motivation on the part of one or more given persons but does not imply that motivation on the part of the describer.

These definitions may be elaborated as follows.

- (1) Appraisals have a first person behavioral significance. They may be reconstructed as answers to the question "What value (significance, meaning, implication) does this state of affairs have for me here now?" My actions are always here now and they are based on the significance of things for me here now. It is because I must act in one way rather than another that I evaluate my circumstances because the results of the evaluation are bases for acting in one way rather than another (recall "assimilation"). For example, if I see that the bear that just walked into the room is a danger to me here now (i.e., that is his relation to me here now) then I have a reason to act to change that relation to one of non-danger. The relation between the cognitive and motivational aspects of this situation is best expressed by an identity statement rather than a

causal, correlational, or merely implicative one: To see the bear as dangerous to me here now is to have a reason (to be motivated) to get out of the danger.

- (2) In contrast, to recognize that the bear is brown has no motivational significance per se, although in some special context the bear being brown might be evaluated in a way which carried motivational significance. "The bear is brown" is a mere description. So also are "This automobile is 3 years old" and "This information system has an average response time of 8 seconds."
- (3) If I am an observer at a distance and I see that the bear is a danger to John and Jenny, my corresponding description "The bear is a danger to John and Jenny" would qualify as a mere description, since as such it implies no motivation on my part--it is simply a fact about the relation of the bear to Jenny and John. However, it involves the same kind of concept ("danger") and judgmental capability as would be involved in seeing that the bear was a danger to me and it implies that it would be appropriate for them to be motivated in the same way as I would be if I saw myself as being in danger from the bear. Because of these basic similarities to appraisal, we distinguish this kind of description as descriptive appraisal.

The logically primary case of evaluation (the paradigm case), or appraisal, is given by the first person, here now, action oriented judgment portrayed in (1) above. The temporary, situational, and person-specific aspects of evaluation have, unfortunately, led to pejorative characterizations of such judgments or phenomena as "subjective" (and therefore to be scientifically avoided--which is itself a case of making an evaluation).

This negative evaluation of the primary case of evaluation is not unfounded. If we want a scientifically or technologically useful generalization, judgments which apply only to certain individuals and which are valid only for brief periods of time in particular places are unlikely to qualify. What has the needed formal characteristics is descriptive appraisal. Descriptive appraisals may attain the necessary portability and durability to make them scientifically or technologically valuable as empirical generalizations and guides for behavior.

Descriptive appraisals can only be adequately understood as derivative from the primary case (the paradigm case) of first person here now appraisals. We obtain the empirical generalizations formally by dropping the first person and here now restrictions. Since evaluation does depend essentially on the action context, generalizations achieved by decontextualizing are at some risk, and it may be a serious risk. Such generalizations cannot be expected to achieve the same degree of rigor as the primary sort of evaluative judgment. I can judge with confidence that the bear is dangerous to me here now. In contrast, if I move to the corresponding generalizations "Bears are dangerous to people" or "Bears are dangerous," that will invite rejoinders such as "Even when they are behind bars?" "Even to the six million dollar man?" "Even when they're asleep?" and so on ad nauseum.

In general, then, descriptive appraisals general enough to be of scientific/technological interest are subject to exceptions and qualifications, so that the empirical/observational basis for a given generalization is of crucial importance. As guides to action such generalizations are subject to qualms and disclaimers (see the "Precaution Paradigm" in Appendix B). An evaluational generalization is accurate (correct, true) because it is valid across the various

contexts of its application, not because we have succeeded in dispensing with "subjective" judgments in favor of "objective" ones.

Because the choice of behaviors has a preferential, or comparative, aspect, the appraisals which serve as the basis for behavioral choice also have a fundamental comparative aspect. "Bears are dangerous" or "Graphic displays are easy to understand" carry relativistic or comparative presuppositions. They invite such questions as, "'Dangerous' as compared to what?" or "'Easy' as compared to what?" The background question is "What would call for any different behavior?" since if everything called for the same behavior we would have no need to distinguish states of affairs and evaluate them.

Simple comparative judgments, e.g., "Graphics displays are easier to understand than tables of 1's and 0's," will be more fundamental than quantitative judgments, e.g., "Graphic displays have a mean rating of 7.6 on the 'Understandability' scale whereas binary displays rate 2.0." This is because they are less decontextualized and because they imply less. One can always make a comparative judgment for any pair of cases. One cannot guarantee that any display which has a quantitative understandability index of 7.6 is always (on a direct comparison) easier to understand than a display which has a lower index. However, to the extent that a numerical scheme succeeds in rationalizing a set of comparative judgments, the gain in utility provided by the quantification may be considerable (since it allows the comparison of two cases which have not been directly compared and in addition the quantification may enter into interesting functional relationships).

Another feature of evaluative judgments which is poorly understood and creates a variety of problems is that many such judgments are double

negatives rather than positive judgments. The notion of double negatives may be explained as follows.

One of the primary forms of evaluation a person makes is a "diagnosis" of what has gone wrong or is going wrong in the current behavioral context. Along with this there generally goes a "prescription" for corrective action. Such diagnoses and prescriptions are classified as belonging to "critic/appraiser" functioning" (see Appendix B). Many familiar locutions and ways of talking are characteristically critic "diagnoses." For example, "Not very convenient to use" and "Output is uninterpretable" are both of this sort.

When a given phenomenon is known to offer the possibilities of going wrong in known ways corresponding to known diagnoses, there is some independent interest, on any given occasion, in the question of whether it has or has not gone wrong in that way on that occasion. Because of this, the critic may appropriately comment on either the presence or the absence of a given deficiency or failure. Thus, the judgment that an information system is "convenient to use" or that "the output is quite interpretable" is likely to represent the judgment that it was not inconvenient to use or that the output was not uninterpretable. Verbally, "convenient" and "interpretable" have the appearance of positive characteristics like "green" or "6 inches tall."

Particularly since such judgments are quantifiable ("degree of convenience," etc.) many people take it that such evaluative judgments refer to some simple property, but then it is an ineffable one. Alternatively, since there is no such property they conclude that this way of talking is merely a cover for an emotional reaction or a recommendation for action. (The discussion of appraisal and the example of danger should suggest why such theories would be plausible.) In fact, however, competent judgments of this kind are

commentaries on what might well have happened but didn't, or on what might happen but in fact doesn't.

Double negative judgments reflect both contextual and comparative aspects of the basic appraisal phenomenon. Once the logical structure and the conceptual and empirical ingredients of such evaluative judgments are recognized, the temptation to operationalize by replacing them with simple observables vanishes. One can then foresee possible technical difficulties, but the methodological aspects are not problematical. (The technique of defining Version Classes described in Appendix E provides a vehicle for dealing with evaluative judgments of this kind.)

APPENDIX E

**A BEHAVIORAL APPROACH TO EVALUATION OF
INFORMATION SYSTEMS**

1. The value of an information system (IS) or of its various characteristics, e.g., speed of operation, reliability, graphic output, etc., is an instrumental or derivative value, not an intrinsic value. It is derived from the value of the behaviors and outcomes which the IS brings about or makes possible. Correspondingly, the differential value of the IS or its features is derived from the behaviors and outcomes which it makes differentially possible.

To say this is to state the context problem in a particular form (see Appendix A, Appendix D), since the behaviors and outcomes which determine the value of the system or its features are external to the system and presuppose a relevant context which is in general independent of the IS.

2. The context problem for information systems can be dealt with by considering both the IS and its context as a single system, more precisely, by identifying an existing system which does serve as the required context. This will be a social system. That is, it will consist essentially of persons who are engaged in structured patterns of behavior which involve, in addition to the people, the material objects involved in those activities.

3. Resources for representing a social system, its elements, and its functioning are presented elsewhere (Ossorio, 1978) and summarized in Appendix B.

4. The value of the information system corresponds to the contribution it makes to the functioning of the social system (SS). The latter can be directly evaluated as more or less desirable. From this evaluation an evaluative index for the information system can be derived.

5. The contribution of the IS to the social system stems from its participation, which in general involves the following.

- (a) Obtaining certain information
- (b) Retaining the information
- (c) Interaction with a human User.

Some paradigmatic forms of interaction are

- (a) Dialogue: Both have some initiative and respond to the other's initiative.
- (b) Master-slave: One instructs and the other executes.
- (c) Consultant: One provides, on request, information which is authoritative within its limits and the other takes into account both the authority and the limits.

6. The evaluation of the IS in this framework is formally identical to the evaluation of one of the human participants in regard to the contributions to the functioning of the social system. At issue are (a) what the individual, P, contributes to the SS, (b) whether that contribution falls within the required (normative) limits of contribution, and (c) which of P's individual characteristics account for the successes and failures.

7. What is required for an evaluation of the contribution of the IS is an evaluation of how the social system functions with the IS as a participant as compared with some alternative.

It would seem that social system functioning can be directly evaluated either categorically or (to some approximation) quantitatively in the sense that its possible (and, of course, actual) historical episodes can be so evaluated (e.g., "satisfactory," "barely acceptable," "disastrous," etc. or "7.5 on a 9-point scale of successfulness").

8. If the social system is represented in the form of a Process/Object, or structure-function description (see Appendix B, Process Description) both actual and possible episodes in the history of the social system can be represented as Versions of the process that comprises a functional cycle of social system functioning. (A Version of a process is simply one of the ways in which the process can take place; when the process takes place on a given occasion that is the same thing as one of its Versions taking place on that occasion.)

9. For most complex processes the number of distinguishable Versions is indefinitely large, so that the set of Versions as such is unwieldy from the standpoint of analysis or description. It is feasible, however, to define a finite number of classes of Versions and to consider the various Versions only insofar as they are members of one or another of these classes, which are designated as Version Classes.

Classes of Versions may be defined in any formally acceptable way, for example, by reference to the inclusion of a given state of affairs, process, event, or object. In general, the Version Class approach provides complete flexibility in segregating the set of Versions meaningfully in terms of formal, process, or outcome criteria or any combinations or elaborations thereof. Two examples are the following.

- (a) A familiar limiting case is that in which the set of Versions is divided into a single class, S, and its complement, U, with S being defined as comprising those Versions having a specified state of affairs, G, in their terminal segment. This will correspond to the case where a desired outcome, or Goal (which is the state of affairs, G), serves as the criterion, with the functioning of the system being appraised as "successful" (S) if that outcome is obtained and unsuccessful (U) if it is not obtained.

- (b) More to the point, we may define a set of Version Classes by reference to a set of scenarios of system functioning which have a direct evaluational relevance. These scenarios would serve either categorically as or Paradigm Cases (see Appendix C, Paradigm Case Formulation) to collect individual Versions into sets. For example various "disaster," "emergency," and "normal" scenarios may be of special interest in assessing the contribution of the IS under a variety of operating conditions.

10. Issues arise and choices can be made concerning how many degrees of freedom to allow in the representation of social system functioning. This is because (a) the functioning of the social system is not a constant and (b) the contribution of any participant, including the IS, will in general be different depending on how the rest of the system is functioning. In this regard it is essential to separate the conceptual approach and the experimental approach.

From the conceptual point of view we would want to assess the contribution of the IS under a wide variety of social system histories, for the same reason, e.g., that we would want to test a carburetor on an automobile under a variety of conditions such as level, straight, normal acceleration; uphill, heavy load acceleration; downhill slalom, and so on. Further, for the sake of quantification and comparison we might be able to define a standard set of scenarios for a whole group of information systems.

However, as noted in Appendix A, social system histories cannot be freely manipulated by an experimenter, and so although we may define Version Classes which have evaluational significance, we cannot in general realize many of these possibilities. One crucial respect in which social system functioning is currently not directly testable is under battle or wartime conditions.

Yet the contribution of the IS under these conditions may be one of the most important considerations, in principle, for evaluation.

Clearly, therefore methodology and procedures for estimating system performance under hypothetical conditions are essential for the most rigorous empirical evaluation of information systems. At the present time only simulation methodology appears to be adequate in principle for this task. (A methodological and procedural simulation model is described in Sections 3.0 and 4.0.)

APPENDIX F
TRADEOFFS

Various aspects of an information system may be considered to be more or less advantageous, other things being equal (and likewise for various aspects of the social system in which it functions). For example, other things being equal, it is more advantageous to have faster retrieval, greater storage capacity, more complete information, and so forth. However, maximization of all such advantages, or desiderata, simultaneously is seldom achievable. Often the various desiderata are inversely related (e.g., amount of information and speed of retrieval), so that simple optimization programs are impossible. Indeed, even when all other desiderata are compatible, there will inevitably be an inverse relation to cost. It is these inverse relationships which create the problems of "tradeoffs" and "cost-benefit analyses."

Traditionally, the tradeoff problem has been seen as one of achieving objectivity and quantification. If only we could quantify the various desiderata rigorously, we could then establish the mathematical relationships which we are crudely approximating with our verbal references to "tradeoffs" and subsequently establish objective points of optimality for various combinations of desiderata.

However, even assuming the relevant quantification (which is presumably available, caveat emptor, on the basis of human judgment scaling) we may expect a strong context effect. That is, in different operational settings, speed of retrieval, more information, etc. would have different absolute and relative values. Because of this any program aimed at quantifying tradeoff functions with a high degree of generality must be regarded as highly speculative rather than oriented toward a practical goal. All of the limitations on empirical generalizations indicated in Appendix D apply here.

The Version Class approach lends itself to systematic tradeoff analysis. In part this is because it is linked to the social system context which allows for a single overall evaluative index. In part it is because the Version Classes may themselves be selected or defined in terms of their relevance for tradeoff analysis.

- (a) Given a single, judgment based, evaluative index for the functioning of the social system, the effects on this index of varying the values of the tradeoff parameters provide a direct means of attempting to optimize those values collectively. They also make it possible to map parametric value ranges which are either safe, decisively risky, or critical division points between safe and dangerous.
- (b) As noted above, one can use a set of scenarios as Paradigm Cases to generate Version Classes which have evaluative significance. That is, the occurrence of a given scenario can be directly evaluated as desirable, acceptable, disastrous, etc. and the effects of information system functioning in increasing the likelihood of one or another of these scenarios is a significant basis for evaluating the information system. Because of this, any given scenario may be used in the same way as described for the overall social system functioning in providing the criterion for evaluating the information system.

Further, it is possible to map the effect of changing parametric values, singly or in combinations, on the likelihood of each of the scenarios individually. This set of functional relationships provides a relatively direct representation of what is being traded for what.

APPENDIX G

GLOSSARY

A-O-C See Actor-Observer-Critic

Actor See Actor-Observer-Critic

Actor-Observer-Critic

One of the basic schemas in Descriptive Psychology is the Actor-Observer-Critic (AOC) Schema. Each of these represents an essential mode of functioning for persons. One version of the Actor-Observer-Critic schema represents these three modes of functioning as sequential and as forming a self-regulating negative feedback loop for the person. The schema generates the basic logic of the Precaution Paradigm for research or evaluation. The three modes of functioning are described briefly as follows.

Actor: In this mode, the person assimilates the world to his own activities.

Behavior is a spontaneous, creative pursuit of the person's own inclinations and the world is viewed as a structure of opportunities and possibilities.

Observer: In this mode the person notes, in a non-evaluative way, what is the case, what it is that is happening, what it is leading to, et cetera.

Critic: In this mode the person evaluates how things are and how they are going and implements the following policies.

- (a) If things are good enough, enjoy and don't interfere.
- (b) If things are not good enough, generate
 - (1) a diagnosis of what is wrong and
 - (2) a prescription for what to do differently as a way of improving things.

The products, i.e., the evaluation, the diagnosis, and the prescription, are the content of feedback to the person as Actor.

Appraisal

An appraisal (evaluation) is defined as a description which logically carries motivational significance for the describer. It contrasts with a "mere description" which carries no motivational significance. Appraisals are person-bound and context-bound.

To appraise an object or state of affairs as dangerous to me here now is to have a reason to act so as to change my circumstances in such a way that I am no longer in danger. To appraise an information system as effective in getting me the information I want is to have a reason to use it.

Appraisal is essential for human behavior, since without it cognition would have no rational connection to motivation or behavior and the self regulation of behavior would be impossible.

C-Space See Classification Space

Classification Space

A completely automatic information system for accomplishing subject matter indexing and retrieval for unformatted text. The system is based on an artificial intelligence model and a psychometric methodology in contrast to the more traditional categorizing, keyword, and concordance approaches. Reference: Ossorio, 1964, 1971.

Configuration Description

The conceptual-notational system for real world representation in Descriptive Psychology includes formats for representing objects, processes, events, and states of affairs, together with certain derivative formats.

The Configuration Description is a case of either (1) an Object Description with formal process components or (2) a Process Description with formal object components. Most real world phenomena which are the object of scientific study are appropriately represented by Configuration Descriptions. Hardware, such as a can opener, an internal combustion engine, or a computer are most naturally represented as objects having process aspects or components. Social systems, behavior patterns, and organizations are most naturally represented as processes having object components.

Because of its general character, Configuration Descriptions are often referred to either as Object/Process descriptions or as Structure/Function descriptions.

Reference: Ossorio, 1978

Critic See Actor-Observer-Critic

Data Independent Accessing Model (DIAM, later designated as DIAM I to distinguish it from later developments)

DIAM is a theory of representing information in physical space developed by the late M. E. Senko of IBM between 1967 and 1978. The theory provides a multi-level abstraction of the process of accessing information. Each level can be viewed as a closed system of operations on an elementary unit. As such, the theory is a comprehensive task analysis of information

management in which each function is allocated one and only one place in the hierarchy. Its elementary units have been shown to be sufficiently general to describe the architecture of all the commonly encountered data base systems to which they have been applied. Publications by Bubenko, CODASYL (SDDTTG), Levin, and Schneider, in addition to those of Senko, show that the theory can be applied to the understanding, analysis, modelling, design, and even the operation of information systems from simple file management programs to complex networks of distributed/heterogeneous data base systems.

Reference: Senko et al., 1972

Data Model

It is common, in the information science community, to distinguish among three related frameworks for representing and organizing information. These frameworks are the Data Model, the External Schema, and the Enterprise Model.

The most familiar of these is the Data Model, which is a representation of the structure of information within the computer or information system as such. A tabular file of Records, and a hierarchical system of Records are familiar generic data models. A model of the data, including relations between different sorts of data is needed in order to identify the data systematically for purposes of storage and retrieval.

The ways in which information is described and identified for (broadly) computational purposes is in general different from the ways it is described and identified for any other purpose, and it is this difference which creates access and interface problems. The most common way to identify information is with a "holistic," or "placeholder," description (Ossorio, 1978).

Phrases such as "my current bank balance," "what's going on in the next room," "what they were arguing about," "the next play of the game," etc., are such descriptions. They are used in asking questions and getting answers:

Q: What is my current bank balance?

A: Your current bank balance is

Q: What's going on in the next room?

A: What's going on in the next room is

Q: What were they arguing about?

A: What they were arguing about is

The concepts involved in these specifications of information are very different from the concepts such as "record," "file structure," "query language," etc. used in a data model. In general such concepts belong to the ordinary world of observation and human interaction, i.e., they are "Real World" concepts.

Information systems are involved in creating real world activities by people, e.g., the conduct of a business, the accomplishment of a mission, et cetera. Each such activity involves a limited range of concepts, terms, decision points, and actions. The function of an information system is to provide the basis for decisions and action at various points in the activity. Information will therefore be identified by giving holistic descriptions employing these concepts, for example "the number of ships within 100 miles of X." The representation of information in those terms in which it has reality within the social enterprise in question is the Enterprise Model.

The External Schema corresponds to a more noncommittal concept of what the information system is used for and how. Once we recognize that the User's activities and purposes which involve the information system are in

general conceptually structured in ways which are radically different from the data model, we may use the term "external schema" to refer to that conceptual structure, whatever it may be. The notion of an external schema does not carry the same implication of coherence and organization that the notion of an Enterprise Model does.

For a current discussion in this area, see Kent, 1980.

Descriptive Appraisal

An "appraisal" is defined as a description which logically carries motivational significance for the describer; a "mere description" is one which does not. Descriptive appraisal is an intermediate case in that it has a third person motivational significance rather than a first person significance. For example:

- (1) "That lion is dangerous to me here now" is an appraisal, because to see the lion as dangerous to me here now is to have a reason to get out of danger.
- (2) "That wall is blue" or "That is a lion" are mere descriptions, since they involve no motivational significance for the describer.
- (3) "That lion is dangerous to those people right now" is a descriptive appraisal. Although it implies nothing about the describer's motivation, it does mention a circumstance which for "those people" would be a reason for action.

The importance of descriptive appraisal is that it provides a link between the basic phenomenon of appraising or evaluating, which is completely context-specific, and the theoretical or empirical generalizations which have scientific or technical value.

A generalization such as "lions are dangerous to people" has some implications for context-specific appraisals and therefore for behaviors. The implication is "If you're dealing with a lion, there is a good chance that it's dangerous to you (so make the appraisal and act on it, or, so don't take chances--treat it as dangerous)." Similarly, "Ease of use is a significant element in successful information systems" is a generalization which has implications such as "If you are dealing with an information system which has a low 'ease of use' index there is a good chance that the IS will not function successfully (so make the appraisal and, e.g., compensate for the deficit, or, so don't take chances--buy the alternative which doesn't have this flaw)."

Reference: Appendix D, Appendix E, this report.

DIAM See Data Independent Accessing Model

Disaster Scenario

The basic criterion for evaluating the contribution of an information system to a social system is provided by classifying the possible histories of the social system into evaluatively ordered classes (in the simplest case, into two classes, i.e., "good" or "desirable" vs "bad" or "undesirable"). A "disaster scenario" is a script-like narrative which depicts an undesirable history. It may be used further as a Paradigm Case to define a class of such histories for purposes of representation, analysis, or data collection.

Double Negative

This term is used to refer to characterizations, such as "efficient," "convenient," "timely," "User friendly," and so on, which, because of their

grammatical use as adjectives or adverbs, appear to refer to some positive and determinate desirable characteristic, comparable to "red" or "sour," when in fact they are used to mark the absence (the first negative) of a certain kind of flaw, deficit, distortion, failure, etc. (the second negative).

The importance of this concept in the present context is that it reminds us that the phenomena of interest are the failures, flaws, et cetera and that assessment properly consists in establishing their presence or absence rather than the (hypothetical) direct measurement of hypothetical properties of "convenience," et cetera. The importance of this feature of evaluation is codified in the Precaution Paradigm.

Enterprise Model See Data Model

Evaluation See Appraisal

External Schema See Data Model

Observer See Actor-Observer-Critic

Object Description

This is a basic form of real world description in Descriptive Psychology. The description is given through the recursive (if needed) use of the Basic Object Unit (BOU). The BOU is a descriptive format which specifies what it is that must be specified about an object or type of object in order to distinguish it from other objects or types of objects. The required specifications include (a) immediate constituent breakdowns into parts which are themselves named

objects, (b) relationships among parts, (c) contingent relationships among parts (this is the vehicle for representing process aspects of objects), (d) membership characters, i.e., the larger wholes of which the object is a component, and (e) equivalent names of the object.

Reference: Ossorio, 1978.

Object/Process Description See Configuration Description

PCF See Paradigm Case Formulation

Paradigm Case See Paradigm Case Formulation

Paradigm Case Formulation (PCF)

This is a much-used conceptual-notational device (see Appendix C) in Descriptive Psychology. The primary function of a PCF is to pick out a range of cases which it is of interest to distinguish from other cases. A major feature of the PCF is that it is applicable on many occasions when a definition cannot be used because the cases of interest have no single necessary and sufficient feature in common.

The PCF is a two-step formulation:

Stage I: Introduce a Paradigm Case.

The description of the Paradigm Case directly identifies some of the cases which are to be picked out.

Stage II: Introduce some number of transformations of the Paradigm Case.

Each transformation is to be read as follows. "Change the Paradigm Case in this way (i.e., the transformation) and the result will still (also) be a

genuine case." A converse approach is also possible: "Change the PC in any other ways except these and the result will still (also) be a genuine case."

In this way, by introducing transformations as needed, we cover the desired range of cases and successfully distinguish these cases from any others. The PCF is particularly useful where it is as important to preserve the differences among the cases as it is to group them together.

Reference: Appendix C, this report.

Parametric Analysis

Parametric Analysis is a basic conceptual-notational device in Descriptive Psychology. A parametric analysis of a given domain or type of phenomenon is given by specifying the ways in which a particular case can be the same as another particular case or different from it. The parametric analysis thus specifies the parameters, or dimensions of variation and in general these will be distinctive for each distinct domain or class of phenomena.

Like the Paradigm Case Formulation, and unlike the Definition, the Parametric Analysis is essentially non-reductive. The particular strength of this device is that it lays the basis for formulating theoretical or empirical regularities which involve the values of the parameters.

Reference: Appendix C, this report.

Precaution Paradigm

This is a basic paradigm for research and for evaluation. It is based on the Actor-Observer-Critic schema as a model of human self-regulation,

paradigmatically on the basis of actual feedback and derivatively on the basis of deduced or anticipated hypothetical feedback.

The formal structure of the precaution paradigm is that there is a course of action (mission, enterprise, etc.) which an Actor intends to engage in. The course of action involves possibilities of failing or going awry, and these possibilities are of sufficient concern to warrant taking some action to obtain a basis for either reducing the concern or taking remedial action.

In principle the Precaution Paradigm involves the following steps.

- (a) The representation of the course of action (mission, enterprise, etc.).
- (b) The representation of the ways of going wrong which are of concern.
- (c) Prioritizing the concerns and, in order of priority, dealing with each concern as follows.
 - (1) Use the representation of ways of going wrong to identify the observations and data which would provide evidence of going wrong in that way or assurance of not going wrong in that way.
 - (2) From among the relevant possible observations, select an accessible subset and acquire the relevant data.
 - (3) Evaluate the data in relation to the concern and act accordingly, e.g., reject the course of action, take remedial measures, continue course of action, etc.

The methodologically basic case involves actually engaging in the course of action and collecting data in vivo. There are limited possibilities of engaging in similar or partial courses of action specifically for the sake of obtaining the precautionary data. These possibilities are the basis for laboratory experiments, bench testing, etc.

Reference: Appendix B, this report.

Process Description

This is a basic form of real world representation in Descriptive Psychology. (The other basic forms are Object, Event, and State of Affairs descriptions). The Process Description is based on (a) a set of tautological rules which define "process" implicitly by its conceptual connections to "object," "event," and "state of affairs," and (b) a parametric analysis which provides a set of specifications which are needed in order to distinguish one process (or type of process) from any other.

The gross part-whole structure of a process is given by specifying Stages and Options. The formal constituents of the process are given by specifying Elements, Individuals, and Eligibilities. The fine structure of the process is given by specifying Attributional Contingencies and/or Cooccurrence Contingencies.

These specifications are codified in the Basic Process Unit, which is a representational unit which can be used recursively (the Options within Stages of a process are themselves processes) to represent a process as a hierarchy of possible histories or "ways things could go." Each distinguishable history of this sort is a Version of the process. A Version of a process is one of the ways in which the process could actually occur, and the occurrence of a process on a given occasion is identical to the occurrence of one of its Versions on that occasion.

Two forms of systematically incomplete Process Description are the Means-Ends Description and the Achievement Analysis, or Task Analysis.

In the present context, Process Description is important as the basis for the Simulation Paradigm for research and for evaluation.

Reference: Appendix B, this report; Ossorio, 1978.

Structure/Function Description See Configuration Description

Simulation Paradigm

This is a basic research paradigm in Descriptive Psychology. It is based on formulating one's knowledge or beliefs, hypotheses, etc. as a Process Representation (q.v.) of the process in question. The paradigmatic test of the adequacy of the representation is to "run" it as a simulation and check predicted outcomes against observation.

A key aspect of such a Process Representation is the accurate specification of functional dependencies which determine how the process goes on a given occasion. These specifications correspond to Contingency statements in the Process Representation. Because of the complexities involving such functional relationships collectively, computer simulation is the natural tool for implementation.

Another key aspect is the high ratio of confirmatory value to amount of data involved. In general, successful prediction provides some degree of confirmation for all of the functional relationships in the representation. It thus provides the basis for empirically grounded confidence in a representation which describes events which are not yet observed but which are of interest in terms of "what would happen if . . .". In turn, this is the basis for criterion level evaluation of the contribution of an information system to the social system it was designed to benefit.

Reference: Appendix B, this report.

Task Analysis

This is an important form of real world representation in Descriptive Psychology. It is a systematically incomplete form of Process Description in that it deals only with the structure of what "gets done" or needs to get done without regard to "how" or "with what."

The result of doing a Task Analysis is a hierarchical structure of achievements (or the corresponding tasks). Procedurally, one may generate a task analysis by going through the following steps.

- A. There is an outcome, or achievement, A , which I want to accomplish. Do I know how to do A ? If "Yes," do it and exit. If "No," proceed.
- B. (The occurrence or existence of) A can be redescribed as (the occurrence or existence of) A_1 and A_2 or $A_3 \dots$
- C. Do I know how to do A_1 , A_2 or A_3, \dots ? If "Yes," do them and exit. If "No," proceed.
- D. A_1 can be redescribed as A_{11} and A_{12} and $A_{13} \dots$
 A_2 can be redescribed as A_{21} or A_{22} and A_{23} and \dots
 A_3 can be redescribed as \dots
- E. Do I know how to do A_{11} , A_{22} , etc.? If "Yes," do them and exit. If "No," proceed.

This is the Identity form of Task analysis (A_1 , A_2 , \dots are jointly identical to A ; A_{11} , A_{12} , \dots are jointly identical to A_1 , and so on). There is a corresponding causal form such that bringing about A_1 , $A_2 \dots$ is a way to bring about A causally. Hybrid forms involve any combination of identity and causal relationships.

In general, large tasks are not something that any person or group knows how to do directly. Rather, they are done, if at all, by doing a large

number of discrete things which somehow add up to the achievement in question. Process representation is a way of codifying the structure of how things come about, rigorously and in as much detail as is known or as is needed. Task Analysis is an elliptical form, but often more convenient or informative, of this codification. It is therefore particularly suited for representing the direct achievements of an information system or its components and how those achievements contribute to the larger achievements (social system functioning) for which the IS was designed.

Reference: Ossorio, 1978.

Version See Process Description

Version Class

The concept of Version Class depends on the prior concepts of Process Description and Version (q.v.). In general, the real world processes which we distinguish and are interested in are complex phenomena whose instances (Versions) are a heterogeneous class. (For example, consider the range of possible instances of "designing a data base," "detecting and reporting a UFO," or "making friends with someone".) A further feature of such processes is that their distinguishable instances, or Versions, are often so numerous that they form an unmanageable collection. Under these conditions it is necessary for practical purposes to reduce the number of Versions dealt with by introducing some principle for grouping sets of Versions into classes of (for our purposes) equivalent Versions and then dealing with these classes rather than with individual Versions. Such a class is called a Version Class.

In the present context, for example, we describe the use of "disaster scenarios" as Paradigm Cases for defining Version Classes of the functioning of the social system in the course of doing precautionary evaluation or of doing simulation studies. Each disaster scenario represents an eventuality which is of independent interest to us (so that the Version Classes are not merely arbitrary) and any Version which resembles a given disaster scenario more than any other falls within the Version Class defined by the scenario.

Reference: Appendix E, this report.