INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Order Number 9305682

Knowledge and tools for quality designs

Schleiffer, Keith E., Ph.D.

University of Illinois at Urbana-Champaign, 1992

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

KNOWLEDGE AND TOOLS FOR QUALITY DESIGNS

BY

KEITH E. SCHLEIFFER

B.S., United States Coast Guard Academy, 1978
M.S., University of Illinois, 1986

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mechanical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1992

Urbana, Illinois

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

THE GRADUATE COLLEGE

APRIL 1992

WE HEREBY RECOMMEND THAT THE THESIS BY

KEITH E. SCHLEIFFER

ENTITLED KNOWLEDGE AND TOOLS FOR QUALITY DESIGNS

BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF DOCTOR OF PHILOSOPHY

DL Marriott
Director of Thesis Research

A. L. Addy
Head of Department

Committee on Final Examination†

DL Marriott
Chairperson

Richard C. Coddington
David H. Offner
Brian G. Thomas

† Required for doctor's degree but not for master's.

O-517

ABSTRACT

The objective of this research has been to develop an understanding of the relationship between the problems an engineer solves while doing design, the knowledge the engineer has available to apply during problem-solving, and the tools which the engineer might use to organize evidence in support of a decision. In more general terms, the research was intended to add to design theory by adding to the description of how engineers accomplish design.

The research incorporated several different approaches. Experience of engineers, including the author, doing design has had a substantial impact on the conclusions drawn. The body of design theory has been reviewed, as well as the literature on problem-solving in other disciplines. Available tools used to support design were evaluated to determine how useable they are and to explore how tools can be used during design.

A quality design is, in part, a design for a quality product. A quality design also incorporates manufacturability, prevention of failures, and other aspects influencing the value of the designed product. Loss of quality of the design can be defined as a cost - to the designer, the manufacturer, the vendor, the user, or society - when the product built from the design fails to meet an intended goal.

Materials selection is an example of a design decision, incorporating both quantitative and qualitative reasoning methods as a selection is made. Materials selection is also typical because frequently the need to push forward to other decisions interferes with reasoning completely through the selection. Manufacturability and failure prevention are examined in detail in the context of materials selection.

The knowledge available to designers, and the forms in which knowledge is recorded, has a substantial impact on how effectively design decisions are made. Methods for reasoning about usefulness of knowledge intended to support design are developed, and two measures of usefulness are introduced: flexibility and coherence of the knowledge.

A clearer understanding on the part of design tool developers of how a tool will be used in design (that is, how engineers do design) will lead to better fit between the design tools and the needs of design decision-makers. The principal paradigm in engineering design is to acknowledge the complexity of the full problem, but to simplify the problem in order to attack it.

Acknowledgements

I am deeply indebted to my advisor, Professor Douglas L. Marriott, who has been a great help during this work. In particular I appreciate his willingness to take risks with me, both in the projects we have undertaken and in originally agreeing to work with me. Professor Frederick A. Leckie has also had an important influence on this work. All of my committee members have been extremely patient and helpful, for which I am quite grateful.

Several other students have made direct and indirect contributions to this project through their experiences in other, related projects. I must mention the contributions of Jess Comer, Steve Bednarz, and John Lichtenfeld.

The facilities of the Battelle Memorial Institute were used in preparation of the final version of this thesis. In addition, many engineers at the Institute have been helpful by sharing their experience and insight.

Last, but certainly not least, I appreciate the support of my wife Debbie, and my daughter Claire, who patiently indulged me for four years; and my daughter Elaine, who has not (yet) asked to sit on my lap while I was writing.

Table of Contents

Chapter

Table of Contents (continued)

List of Tables

Table

List of Figures

Figure

Chapter 1
Research Methodology

The research reported in this thesis is concerned with how engineers design, in particular how knowledge is employed when solving design problems and how design tools are - and might be - employed when solving design problems. There is a growing interest in design theory research, with substantial emphasis on descriptions of how engineers should design in some idealized setting, or how computers could design in certain limited circumstances. In contrast, the main thrust of this research is to make a connection to design problem-solving as it actually takes place in the messy, complicated real world.

The first chapter is devoted to a description of how this research was carried out. Terms will receive definitions, the working hypothesis and objectives of this research will be given, and the methods used in the research will be described. The second chapter, continuing development of the background, contains expanded definitions of design and introduces key concepts underlying the design theory developed in this research and presented in this thesis, in particular the two concepts of design quality and engineering productivity. In the third chapter, the author reviews the history and current state of design theory and design research.

The fourth and fifth chapters contain descriptions of two experiments into how engineering knowledge and design tools can be used to support design decision-making. The first experiment, described in Chapter 4, deals with better anticipating manufacturability while selecting materials of construction. The second experiment has to do with anticipating possible sources of material failures during material selection, and is covered in Chapter 5. Both of these experiments address the question of catching and correcting potential errors early, when the cost of correction remains small.

In chapter 6, a simple method is presented for visualizing how well the knowledge coded into a tool will meet the needs of an engineer making design decisions. A theory is developed of how knowledge is used for design, and how engineering knowledge can be made more available for design problem-solving. In chapter 7, this theory is applied in an extended discussion of computer aids to engineering.

The eighth chapter is devoted to a broad discussion of the philosophy underlying engineering design, intended as food for thought to one who might wish to pursue design theory research. The ninth and final chapter contains a summary of conclusions and recommendations.

1.1. Working Definitions

One of the problems encountered by design researchers is the expressions used to describe aspects of engineering have varying meanings. In this section, the author will present definitions of terms as they are used in this thesis. No claim is made that the definitions given here are canonical, or in any way "more" correct than another's usage; the author is simply presenting the meanings he attaches to the terms, based as much on experience and personal preference as on any research.

Some of the expression defined here will receive additional consideration later. Chapter 2 will be an extended discussion of many of the terms related to design and problem-solving. Expressions related to knowledge, such as flexibility and coherence of knowledge, will be given working definitions now, and will be more fully defined and discussed in Chapter 6.

Design theory is the body of engineering knowledge which attempts to define and explain design. Two categories can be used to classify uses of design theory. First is to explain the methodology of design to those who actually do design products, with the intent that better understanding of the process will make them better at design. Second is to explain design to those who produce things - new tools or new technologies - which are intended to be delivered to engineers for application. The first category is how to design; the second is telling non-designers how design is accomplished and how to support the design activity.

Design is both a verb and a noun. The verb form refers to the activity of design, the noun form refers to the design as an entity, the result of the design activity. Each of these forms requires discussion and introduction of related expressions.

Design, the verb, is the action of developing and describing a solution to a problem. Design is synonymous with product development; conceiving and developing a device which will solve a problem. An example of the kind of problem to be solved during design is material selection, which will be discussed in detail later in this chapter. Because the author's experience is in the area of mechanical design, it is assumed the solution will be built as hardware. However, the discussion of design, and the conclusions to be drawn about design tools, are generally applicable to design within any engineering discipline.

Each engineering decision is intended to advance the design, that is, advance the description of the solution to the problem. Advancement of the solution can take place in two ways: the overall structure can be expanded to incorporate an additional feature, or the description on a feature can be defined in greater detail. Both a broad view of the complete design, and a close-up view of the component currently being developed are necessary for design to succeed. Decisions (and descriptions of the design-under-development) which deal with the overall picture are called global in this research. Decisions dealing with individual components are called local.

The design, the noun, is the description of the solution to a problem. Again, because of the author's inclination to mechanical design, it is assumed in this thesis the solution takes the form of a hardware product. The description contains sufficient information to fabricate and operate the product which will solve the problem. The design is a template from which hardware will be built, and errors in the template are passed on to each unit. An error in a single production unit may always be insignificant (at least statistically; it is significant to the customer who buys the faulty unit). An error in the design will always be significant because it is reproduced in every production unit, and for that reason the engineer strives for an error-free design.

Each design is produced in response to some form of contract, formal or informal. A contract may take the form of a formal specification-and-proposal document. A contract may take the form of an agreement to the amount of labor and material costs to be expended in search of the solution. Even for an independent development project, the engineer sets goals and expectations for himself. Regardless of the formality of the contract, there is an expectation the solution will take shape within the bounds of the engineer's original estimate of costs and schedule. Each design has a client, possibly internal to the organization, possibly an external customer. For projects which have been broken down into smaller divisions, the client may be another engineer working to coordinate the entire program. There is always a client whose expectations and requirements must be met.

There are many constraints which govern the design process. The contract places before the engineer some requirements which must be met. The problem requires exploration and must be better understood before it can be fully solved. The problem situation may very well change as the design is developed, making it necessary to monitor the clients' needs. Compliance with codes and standards required by law may apply additional requirements. The engineer is given a budget and a schedule to meet. The technology available to apply to the problem is continually advancing. In short, the engineer faces considerable constraints, and those constraints can change with the passage of time.

Implicit in the definition of design given above are the concepts of problems and problem-solving. The problem to be solved is typically a situation which has undesirable aspects; some or all of the undesirable aspects are identified and targeted by the client. The problem is normally stated as a collection of desirable goals to be met combined with undesirable goals to be avoided. The problem statement provided by the client is almost always incomplete; sometimes because the client is willing to accept a solution addressing only a portion of the problem, and other times because the client is not in a position to state the problem more fully. Ordinarily the overall problem is broken down into a set of subproblems to be addressed individually; not because the subproblems are really isolatable, rather because experience has shown the subproblems easier to solve in isolation, even though an additional burden of coordinating the interactions between subproblems is added. This breakdown might be carried out through several levels of hierarchy. Thus the problem to be addressed by the engineer may be the complete needs of the client, such as a new earth-mover better suited to clay soils, or it may be a subproblem incorporated within the contract, for example the choice of material for a power take-off shaft on the earthmover.

Problem-solving is a blanket term used to describe many different activities. In engineering design, problem-solving means making decisions which advance the progress of the development of the complete design. Experience plays an important role in streamlining the problem-solving process. "Good practice" and "sound judgement" are derived from experience, permitting the engineer to skip past certain issues to focus on the ones which matter in a particular instance. A principal reason new and emerging technologies are more expensive to implement is the lack of experience, and resulting lack of rules defining what is good practice for that area.

When the product is built from the design, its value will be evaluated. Two measures of value were considered in this research: performance and quality. Performance refers to the product meeting the goals which were set for the design. Quality refers to the product maintaining is function over time; loss of function equates to loss of quality because the product is not always available to perform as required. When applied to a product, these expressions have a tangible meaning. To discuss engineering productivity, it is necessary to consider performance and quality of the design itself.

There are a lot of factors driving a heightened interest in productivity of engineers. Corporate managers believe that moving the product faster from concept to market introduction enhances the organization's ability to compete. The increasingly competitive marketplace is causing greater emphasis on reducing the costs of developing and producing products. Both the time and cost constraints are motivation to limit the amount of prototyping before the design is released for production.

The cost of developing a product could be evaluated in terms of efficiency: cost and time required to produce a new design. For the purposes of reasoning about design this can be reduced to an efficiency measure for each design decision: cost and time to make each decision, which we can call performance. But how does one evaluate the effect of a design decision on efficiency of production and marketing? There is no obvious connection between efficiency of decision-making and whether those decisions have an impact on efficiency of fabrication and assembly. The quality of the design is determined by whether the fabrication process is as easy as it can be; as well as other considerations such as reliability (quality) of the product itself, or whether it is easy to operate by the intended user.

A design decision must be not only efficiently made, it must also be effective. A second measure of engineering productivity is necessary because engineers are obliged not only to use resources wisely when making decisions, they are also required to be correct in their decisions and to be sure their decisions allow wise use of resources during all of the product's life.

The most obvious example of an effective decision is found when the decision is not in error, and need not later be corrected. But there are other effective decisions as well. A subsystem which has been coordinated to fit well into the overall system is the result of effective decisions. A decision which enhances manufacturability of the product instead of hindering it is an effective one. Decisions which minimize risk of an unexpected failure, or decisions which improve reliability are effective, as is a decision which enhances maintenance access to critical parts.

Productivity of engineers must be measured by considering both how efficient they are at making design decisions, and how effective those decisions prove to be. A balanced combination of both efficient decision-making and effective decision-making is important in engineering design. When considering productivity of engineers, it is important to consider both the effectiveness and the efficiency with which designs are produced.

Efficient decision-making implies the decision was made with economy, or with minimal expenditure of resources. Economy of problem-solving is driven by the way the engineer reasons through the problem, which is largely influenced by the availability and form of the information required to understand and act on the problem. Economical problem-solving is equivalent to efficient problem-solving.

Ordinarily, the most efficient decision is the one accomplished in the least amount of time. Time is important during engineering design for two reasons: time is associated with labor hours, a direct cost of development; and time on the clock for each decision sums into time on the calendar required to turn an idea into a marketable product. Because economical decision-making is important in engineering, this research has been focused on problem-solving which requires minimal reasoning. Minimal reasoning does not necessarily mean the reasoning, written as a syllogism or a logical proof, has a small number of steps, it only means the solution needed at the moment was obtained with minimal expenditure of time.

Minimal knowledge is the knowledge which supports efficient, economical reasoning. It is possible to state the knowledge about a particular design problem type in many ways. A structural problem might be solved with simple statics, by strength of materials methods, through the theory of elasticity, by application of energy principles, or by a numerical simulation such as the finite element method. Each of these methods is a restatement of the underlying mechanics of solids. Depending on the nature of the situation - for instance, a determinate versus indeterminate structure - one or the other may be the basis for minimal reasoning about a structural problem.

In some cases, minimal reasoning is accomplished through the use of other resources, such as a computer workstation. Because the workstation is generally an overhead cost, or amortized over several projects, it is assumed by the author the cost of the computer itself is not a factor in evaluating engineering productivity. Thus a computer can contribute to efficient decision-making, for instance, if it contributes to the decision being made more rapidly. It is also necessary to ask if the decision is made any more or less effective when the computer is included in the process.

A design tool is any technique, method, or mechanism which is intended to support design decision-making. Computer-aided engineering (CAE) is the subset of design tools which require a computer in order to be used. Design tools may be built with emphasis on improving efficiency of decision-making, or on effectiveness, or both. Design tools may be built to deal with a specific, narrowly defined problem, or to provide broad support to a variety of type of decisions.

Design tools represent ways to encode knowledge in a way that improves the progress of design decision-making. For that reason, a study of how design tools may be improved requires consideration of the knowledge needed for design.

A knowledge base is the body of knowledge available to support the activities of a discipline. One might imagine the knowledge available to support engineering design is the entire contents of the library. However, much of what is recorded in the library's holdings is not of direct use during design, and there is a significant portion of knowledge for design which has not been recorded in a form suitable for the library. In the latter category are data specific to classes of parts, information which is proprietary and closely controlled, general rules of good practice, as well as product literature.

Knowledge is not a single, monolithic mass of information. Knowledge has structures imposed on it, in part by the way we choose to organize it, and in part by the media we use to record it. In many cases more than one structure can be used, and a researcher or an author is obliged to choose a structure best suited to the intended audience. Typical types of knowledge structure include tables and charts, figures and illustrations, and different forms of text descriptions. Within each of these types, there are a large number of variations.

The expression data structure has a more specific meaning, normally in the context of computer science. A data structure is a detailed method of organizing data in a way that clearly defines how it may be recorded, retrieved and read, or altered. Data structures have an impact on the discussion of knowledge structures because the use of a computer requires a data structure, and computers are a dominant force in modern design tools.

Knowledge has attributes which influence how well it can be used for design. For the purposes of discussion, the author will now state brief, working definitions of two knowledge attributes: coherence and flexibility. These terms will be developed more completely in Chapter 6 and applied to describe in concrete terms how knowledge should be organized for use in engineering design.

Flexibility means the knowledge base can be adapted to a variety of problem types. Flexible knowledge is capable of supporting reasoning about a problem at different points in the design process, so that the same knowledge base can be applied to deal with both the global and the local aspects of the problem. This assures that one decision is not made in isolation from the rest of the decisions which together create a product. The different ways to approach evaluation of a structure, discussed above, provide an example of flexibility because different kinds of problem-solving methods, suitable for different kinds of problems, are all available. Flexibility is achieved when the knowledge is spread out across a wide variety of potential applications.

If flexibility is how widely the knowledge base is spread, coherence is how well-connected it is. Coherent knowledge has connections between different approaches to a problem type, providing clues about the connections between the knowledge base and potential applications. This assures that a global decision can be made in the context of later local decision needs. The load and stress analysis methods already mentioned form a coherent subset of the knowledge base because the connections among the different areas are visible, and the reasons for using one approach over another are relatively clear. Manufacturability has been a troubling issue in recent years because it has been allowed to lose coherence, and became poorly connected to the main body of knowledge used by engineers for design.

One additional set of expressions requires definition: material of construction and material selection. The material of construction is the material chosen by the engineer to be used for fabrication of the component or product. In some cases, toward the beginning of the development of the design concept, the material of construction is only vaguely defined, for instance in the choice between stainless steel and aluminum for the skin of an aircraft. Broad choice of material is a decision normally made early in the process of developing the design.

In a more detailed decision, the material could be a specific alloy type, for instance 4340 steel, oil quenched and tempered to achieve a compromise between high strength and fatigue resistance for a suspension knuckle. In a limited number of cases, the material of construction is a driving factor in the entire design, such as a 9% chromium, 1% molybdenum steel alloy for steam generator tube platens in a power plant which had to be ordered years in advance.

Material selection is the process of advancing the choice of material of construction. Early in the design process, the material may be chosen in broad terms, such as in the generic decision to use an aluminum in a weight-critical assembly or structure. As development proceeds, a specific alloy class will be chosen. Later, material selection must focus on deciding the particular material to specify for fabrication on a part drawing. Material selection will be the subject of a discussion in a section later in this chapter.

The definitions given in this section provide a basis for thinking about design. Some of the terms common to design theory are ambiguous in meaning and the author has made an effort to avoid using those terms. Where the terms seemed unavoidable, an explicit explanation of their meaning has been given for the purposes of the discussion to follow. The concepts defined above will reappear multiple times in later chapters. Additional terms will be introduced and explained in the course of particular discussions.

1.2. Working Hypothesis and Research Objectives

The objectives of this research are:

- Define the concept of quality of the design.
- Expand the concept of engineering productivity to incorporate both efficiency and effectiveness of decision-making.
- Clarify the nature of the knowledge used by engineers for design, with particular emphasis on improving engineering productivity by changes in the way knowledge is made available.
- Draw some general conclusions about the design of design tools, with particular emphasis on computer aids to engineering.
- Demonstrate approaches to design which illustrate and apply the author's ideas.

- Demonstrate applications of design problem-solving which address global decisions, and those taking place early in the development of the design.

The author has several years' experience as a practicing engineer, including a few years doing design, which have led to a collection of personal observations about design and the support provided to design by other sources. From these observations, the author has constructed a working hypothesis and set the research objectives.

The author's observations about design may be summarized in the form of a set of comments on the present condition of the engineering knowledge base. In general, the knowledge available to engineers:

- Lacks flexibility in that it is not adaptable to new or unusual situations.
- Lacks coherence in that different versions of the same knowledge are not obviously connected together.
- Supports analysis of completed designs but does not support predictions in the course of creating a design.
- Provides considerably less support to early, global decisions than to detailed, local decisions, made very late in the development of the design.
- Does not provide adequate economy in reasoning through design decisions.

The author's observations can be restated as questions for research:

- How can the flexibility of engineering knowledge be improved?
- How can the coherence of engineering knowledge be improved?

- How can the predictive value of engineering knowledge be improved?
- Can knowledge used to support local, detailed decisions be "moved forward" to be useful during early, global decisions?
- How can economy of decision-making be improved?
- Is heavy dependence on the computer necessary or desirable in design? What should be the role of the computer, and of other bases for design tools?

All of these questions deal primarily with the question of what knowledge is provided to engineers, in what forms, and through what media. The answer to these questions is summed up in the working hypothesis of this research:

> The principal paradigm of engineering design is to acknowledge the complexity of the whole problem, but to work on the least complex simplification of the problem for which the resulting approximate solution still seems valid in the context of the whole problem.

Stated in another way, engineering design is accomplished through minimal reasoning. In some cases, where the design process has an iterative nature, a decision may be revisited again and again. A structure to house some field equipment will be evaluated roughly for strength, probably by a quick hand calculation based on beam theory, for the principal loads it will see. Later, a more detailed model might be used for the particular problem of a joint in the structure, and a number of expected load cases evaluated. As hand-holes and other openings in the structure are identified, their effect on the structure will be assessed. Very late in the detailing, the structure might be re-evaluated again for the special case of lifting and handling during installation. This was the sequence the author used when building a structure housing field equipment which is currently in service. Each of the problems associated with developing the structure will be handled by minimal reasoning, with the context of the problem and the solution needed helping to define what method is appropriate for achieving the required economy.

Of particular interest is the influence on design theory of the ready availability of the computer. Many computer aids to engineering design have been introduced in the marketplace, and still more have been developed in research settings. In the author's experience, the tools available to support design, with or without a computer, were developed on the assumption that all the information will be available when the problem is to be solved. The "machine design" textbooks of Shigley and Mitchell (1983) or Juvinall (1983), focused on analysis of a design and not its creation, are examples. The main focus of the research described in this thesis is on methods for proceeding with the design in the absence of complete information.

1.3. Approach

There is no single, universally valid view of what constitutes design, because one's idea about what constitutes design depends to a large extent on personal experience in doing design problems. The author's understanding of design has developed from five main sources.

- Personal experience at engineering design.
- The literature on engineering design theory.

- Interviews and discussions with engineers engaged in product development and manufacturing from a wide range of disciplinary specialties, working in a variety of settings.
- Developing design tools for specific needs, and using or evaluating several prototype and commercially available design tools.
- The general literature on decision-making and problem-solving, taken from a wide variety of disciplines, including psychology, urban planning, management, and mathematics education.

While design theory encompasses a wide range of well-established principles, personal experience is vital to understanding design. Problems requiring complete, "cradle-to-grave" involvement in the project are especially useful because the day-to-day work of developing a design is considered, as it must be, in the context of the whole problem of producing a product, meeting required specifications, and meeting time and budget constraints. Such a project was the author's first experience with manufacturability and assembly planning, in which not only the part drawings but also much of the planning toward inexpensive fabrication were responsibilities, as was actually assembling and testing the system. The engineering firm and the project were both small enough that it was possible for an individual to be involved in the complete program.

In the past few years, the author has enjoyed professional practice with design experience related directly to issues considered in this thesis. This experience includes material selection, drawing checking and design checking (considered in detail in Chapter 5), accounting for manufacturability while detailing a design to be manufactured elsewhere (the subject of the fourth chapter), and test and evaluation of computer-aided engineering tools being considered for purchase, which receive commentary in Chapter 7. In addition, the author has led customers who were not engineers through the product development process, explaining the purposes of and need for each step being proposed.

It is useful to recognize the limits of experience. The author has dealt primarily with projects where single units were to be built, or at most about ten units. This has insulated him from some of the concerns found in a mass production industry, where designing to take advantage of economies of scale during manufacture has a larger role. However, the author's experience with "one-of-a-kind" designs has emphasized maintainability to an extent greater than many engineers in industry. This illustrates how differing experience can lead to differing views of design.

From the author's viewpoint, experience of working through design problems is the best basis from which to approach research about design. John Gero (quoted in Latombe, 1978) commented that many researchers who are engaged in CAE development do not recognize and benefit from reasoning about the design problems (typically software design) which are part of design tool development.

Gellatly (1986) discusses how the early development of cognitive psychology depended heavily on introspection, reasoning about problem-solving by watching oneself solve a problem. Introspection provided the bridge between hypotheses which could be tested by experiment and abstract notions of psychological processes. Design theory has a large supply of abstractions;

introspection into the design process will help to connect those abstractions to the ways engineers really work.

Alexander's (1968) book was originally written on a theoretical, academic basis. Alexander then spent a year in practice, attempting to apply his design theory for architecture while doing architectural design. With the publication of the paperback edition, his preface indicates his insights and methods were quite different after that year. Alexander did not retract his formulation of design theory, but he did note that with experience he had found approaches to the same problems which could not be expressed as easily as the original theory, but which appeared to be more successful. He also noted the original formulation of his design theory did not account well for the needs and wishes of the people who would use the buildings designed. In general it can be observed that the experience of doing design work provides insight and, perhaps more important, a pragmatic test of theory. Without that insight the ideas developed as theory run a greater risk of having limited useful application.

It is not universally valid, however, to say insight cannot be obtained from other sources as well. Asimow's (1962) book, Nadler's (1981) book, and the recent monograph from Koen (1985) feature considerable insight into engineering design despite the authors' predominantly academic backgrounds.

Comer's (1987) interviews with engineers in industry have been useful in consolidating individual experience into a collective conclusion. Although Comer's interviews were conducted in the context of a research problem on education of engineers, the discussions dealt extensively with design practice in industry. Reviews of Comer's transcripts have been supplemented by follow-up discussions with some of Comer's subjects as well as discussions with engineers at the author's current employer. The ideas expressed in this thesis are well supported by the experience of engineers in industry. At several locations in the thesis, examples drawn from these discussions will serve to illustrate the points made. Where evidence from interviews is cited, no names are given because of the anonymity of the interview recordings; the origin of the quote is given by a combination of job description and type of industry.

Nowadays, design tools usually take the form of computer software. Begg (1984) notes that observing the interaction between designer and tool provides insight into the nature of design by allowing an external view of the engineer's internal reasoning. Further development of computer aids to engineering represent a majority of current design research. Certain types of problems are supported by the tools currently available, while others are not. The author has evaluated several design tools in the past few years, both as part of this research program and as part of his employment.

Some of the experimental exercises reported later in this thesis were developed as design tools. Design tool development is a useful technique for testing an aspect of design theory, because a CAE tool can be distributed and tested rather easily. The author is currently using in professional practice the tools developed for this research.

Authors in a variety of other fields have made comments about problem-solving which are of interest to engineering design. The study of decision-making originated in economics, with a good decision defined in terms of profit expected from the outcome. The developing field of management has a heavy emphasis on problem-solving. There are several disciplines, not normally considered directly connected to engineering, which place a heavy emphasis on design, including architecture, urban planning, and industrial design, and there is a body of useful commentary in the literature of these fields. During the last decade, mathematics educators have developed a strong interest in teaching problem-solving as an underlying principle, and the literature in this area is worth reviewing when studying problem-solving in engineering.

The foundation for this research is quite broad. As illustrated in Figure 1, the subject areas considered form a nested set of topics, each one applying to engineering design. The breadth in this research has established a clearer insight into the particular problems of engineering design, and also has shown how design has many characteristics in common with problem-solving in other disciplines.

1.4. Material Selection

In artificial intelligence, it is customary to exercise a new idea through a "toy problem". A toy problem is one which is complex enough to challenge the problem-solving technique, but still simple enough to be comprehensible and reasonably limited in scope. In much engineering research, it is also the practice to focus on a particular example problem while developing theory and methods of broader intended application.

In this research, the author has focused on the problem of choosing materials of construction as an example design problem. Material selection is not a trivial issue in design. Yet material selection is a reasonably constrained issue, and therefore suited to a research program. In addition, material selection has not received much attention in design research to date.

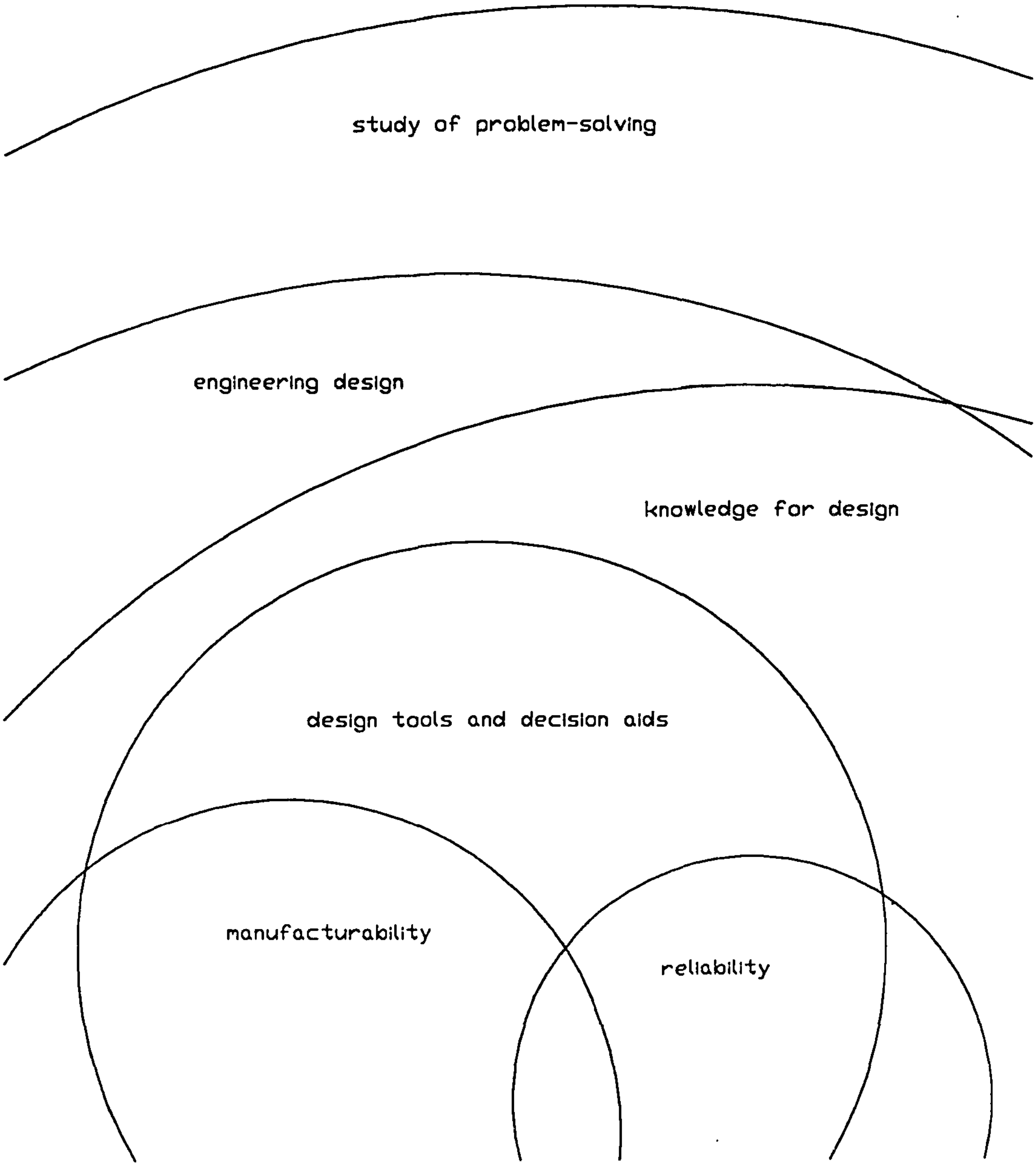


Figure 1

Problem Areas Addressed and Their Interactions

Material selection is a complicated problem, reflective of most other design problems. Although one might wish an optimal choice of material, optimization in the formal sense is difficult, if not meaningless, because of the nature of the material selection process. Several factors contribute to resistance to optimization methods: the choices are discrete, rather than on a continuum and are irregular in their distribution; several material properties and design goals compete as selection criteria; it is usually necessary to compromise between conflicting constraints; and the objective function would be difficult to formulate and evaluate, and would require reformulation when addressing a new selection problem.

Material selection is a typical example of a design decision for precisely the reasons given above. In general, design decisions are between discrete options - even the "optimal" solution of the proper thickness for a pressure vessel shell is driven by available standard plate thicknesses. Decisions are driven greatly by the context of the situation, rather than by any universal standards of what constitutes a good choice. It is frequently necessary to compromise, settling for the best that can be accomplished rather than the exact letter of the requirements. Enhancing one feature of the design is usually accomplished only at the expense of another feature. The definition of the "best" choice is a cloudy one, normally settled by judgement rather than by an objective assessment.

As an engineer designs, many ambiguous and potentially conflicting goals must be balanced. A good example of the conflict inherent in design is the set of compromises required when selecting a material. There are many trade-offs, such as that between ductility and strength in a metal or between obtaining superior properties, but increased cost, through alloying. In some cases alloys have been designed to suit particular applications; such as the recent high-strength, low-alloy (HSLA) steels and the nickel superalloys. Even where design of particular material properties to meet the needs of specific design applications may become possible, such as with composite materials, the engineer still needs a broad notion of the ranges of properties obtainable in order to specify the custom material design. The engineer also requires an understanding of all the material properties, even those which aren't amenable to tuning.

In most designs, developing a material specifically for the application is not a viable alternative. It is necessary to identify the governing material properties, search available materials for candidates meeting those criteria, and then check that other properties, not included in the original search criteria, are acceptable. Depending on the type of problem, the search may concentrate on materials with the required mechanical properties. Yet the material must be modified into the geometry specified in the design, and material properties can have a strong influence on ease of fabrication. The material carries with it the potential for certain forms of mechanical failure, which must be prevented to preserve proper function of the part. Manufacturability and failure prevention are two examples of selection criteria which are often treated as secondary.

We ordinarily assume that some portions of the decision, such as consideration of loading and strength, can be evaluated through analytical means. For strength in monotonic and uniaxial tension such an evaluation is possible, but for most loadings of genuine interest - such as multiaxial, varying-amplitude cyclic loadings - there is no clear approach which will serve all needs. The generally accepted approaches to such a stress calculation are detailed and therefore expensive. Compounding the difficulty, there are many considerations in material selection which are not readily evaluated, including many of the less-familiar failure modes as well as much of manufacturability.

Material selection can proceed in the following manner. There are usually one or two primary considerations governing selection. In the primary assessment the engineer separates the few potentially useful candidates from the remaining, many materials which are obviously not suitable candidates.

In some cases selection is governed by a mechanical property:

- strength
- wear resistance
- fracture toughness
- creep resistance

Selection may be dominated by economic concerns in certain decisions:

- cost
- available forms
- manufacturability

There are a wide range of physical and chemical properties which may be considered during primary selection when the working environment is unusual, such as:

- corrosion or oxidation resistance
- thermal and electrical properties
- tendency for embrittlement mechanisms
- stability of the microstructure and properties under thermal, radiation, or chemical effects
- effects of the material on the environment

The purpose of the primary selection step is to reduce a large list of all possible materials to a small list - perhaps two to twenty - of materials which will fill the main functional need. At that point, the engineer could stop with an arbitrary selection (or an educated guess) of one material from the candidates, guided by judgement and experience. In many situations a choice made at this point is good enough, since the other concerns, not explicitly considered, are of little enough importance that their effect need not be considered. However, there have been cases where a concern which was not considered during selection led to such problems as unexpected difficulty in manufacturing, or an unanticipated mode of material failure leading to poor reliability of the designed component.

The next step in material selection is to proceed with a secondary evaluation, considering those properties which were not included in the primary screening. The goal in this secondary screening is to consider all other possible concerns. The problem is that there are many secondary considerations, and it is rarely feasible to consider all of them. Thus at some point material selection is a decision made by judgement, based on available evidence which is known to be incomplete.

Engineers with experience, or working in a project which has a firm foundation of prior experience with similar products, have an advantage. The material best suited to the component has developed by evolution, through time and experience, and for that reason there is a sound basis for simply re-using a material which was used in the past for a similar application. It is still necessary to consider what makes the new design different and how that difference affects the material choice. In contrast, there is no such basis when the material being considered is a recent introduction. This illustrates the greater degree of uncertainty, and therefore risk, which is present in an innovative solution.

The distinction between primary and secondary evaluation is that, historically, secondary evaluation was not considered explicitly, although in most cases it was dealt with as a portion of the experience applied to the decision. In some cases, the secondary concerns which were not explicitly addressed had an effect on the quality of the component. Quality of a designed item, therefore, can be improved if more of the secondary concerns of material selection can be handled. The problem is to do so without unnecessary additional time or effort on the part of the engineer.

Very little of material selection is driven by the particular details of the component because most of the details are not fully defined when the material of construction must be fixed. The material should be chosen based on a number of criteria, including not only an anticipation of how it will perform, but also its availability, its cost, its ease of fabrication, and its tendency to other failure modes. In common practice, some but not all of these issues are considered during material selection. To be able to incorporate more of these concerns into a material decision, the engineer needs knowledge about them to be stated in forms which fit the ways design decisions are made.

1.5. Summary

Design is the process of developing a description of a hardware solution to a problem. The design process is constrained by a wide variety of requirements, including technical specifications to be met, budget and schedule to be followed, and accounting, during design, for the remainder of the product's life-cycle.

The ways knowledge is structured when it is recorded influences how easily it can be used for design. The knowledge base needs to be flexible to accommodate the wide variety of decisions which must be made. The knowledge base must be coherent to assure each decision is made in the context of the entire development program.

Productivity of engineers must be measured by considering two aspects of decision-making. Efficient decisions are obtained with minimal expenditure of resources; the primary resource used is time. Effective decisions are those which contribute to a better design as a whole.

Engineers accomplish design by minimal reasoning: simplifying the problem until a solution which meets the immediate needs of a decision is obtainable. Minimal reasoning helps efficiency by allowing the assessment underlying a decision to be carried out quickly. Minimal reasoning can assist effectiveness by supporting consideration of a broad range of issues while preserving economy of the decision.

Material selection is a typical example of a design decision. There is a nominally large set of possibilities which can be narrowed by a fairly simple screening operation. Properly made, the final selection requires consideration of more constraints than are convenient to handle, and some judgement is required. It is rare when there is a single right answer for the problem, often several options appear equally workable. Care is necessary to avoid overconstraining the choices to the point no options remain.

A missing element in the tools and knowledge available to support an engineer engaged in design is support for the need to proceed with decisions in the absence of full information. The remaining chapters of this thesis will be focused primarily on that problem.

Chapter 2
Design and Productivity

In this chapter, further consideration of design will expand on the working definition sketched out in Chapter 1. Additional detail and discussion is necessary to clarify what makes design a difficult problem in engineering. Following that will be an examination of the reasoning methods used to solve problems in engineering design.

The discussion will then turn to the meaning of quality. Product quality is now a fairly well-understood concept, one which even has rigorous definitions to permit objective assessment. The concept of product quality can be extended to the design, which is the product of an engineering decision-making process. A design can also have good or poor quality associated with it, and quality of the design has substantial implications for the success of the product defined in that design.

At the end of this chapter, the productivity of engineers engaged in product development is considered. Economical or efficient problem-solving is important to continued prosperity in an increasingly competitive market. Equally important, and less well-understood, is the need for design decisions to be made effectively. The distinction between efficient and effective decision-making was made briefly in the definitions of Chapter 1, and will receive a more thorough discussion in this chapter.

2.1. On Design

In the first chapter, design was described as the process of making decisions which advance the description of a solution to a problem. Overall the process was called product development. In this section, some of the implications of that definition will be explored. More detail will be developed about the requirements and constraints the design must meet. Some of the properties of a design project with respect to time sequence will be examined.

Design problem-solving is characterized by the search for a solution to a problem. The solution takes the form of a plan or description which explains how to make and operate the actual solution. A prototype may be built, but the prototype does not solve the problem, it only provides a demonstration that a critical portion of the design is in fact going to work.

Each decision made must advance the design, that is, it must advance the description of the solution to the problem. All decisions are subject to constraints. The constraints may be external to the design group:

- explicit requirements in the contract.
- explicit requirements of the law.
- an established budget of costs to complete the design.
- an established schedule of milestones for completion of significant portions of the design, and a program completion date.

The constraints may be internal to the design group:

- The need to coordinate development of separate subsystems of the design.
- The need to coordinate the decision-making within several different disciplines developing the design.

- The need to coordinate the decision-making of several individuals, even within a group focused on a single aspect of the design.
- Long-term aspirations of the designing organization, such as market posture, market expansion, and prestige.

The constraints can be classified into two categories: technical requirements, and management goals. In general, technical requirements will take priority. A client is much more likely to accept and accommodate a cost overrun than a failure indicating the project will not result in a workable, marketable product. The discussion to follow, and in general this research, focuses on technical requirements.

Technical requirements, in turn, can be divided into two categories: those describing positive goals to be achieved, and those giving negative goals to be avoided. Positive goals to be achieved normally deal with the performance of the product, and are, for this research, called performance requirements. Negative goals to be avoided normally deal with the quality of the product and are, for this research, called quality requirements.

Most performance requirements can be stated explicitly. Performance goals for mechanical systems might include the rate at which work is to be done, such as in moving fluid or a mechanical load; or the velocity and acceleration ranges for a vehicle. Performance goals for a mechanical component might be a family of cyclic loadings and an intended service life for an engine connecting rod, or a minimum leakage pressure for a valve.

Some quality requirements can be stated explicitly, but a majority are stated implicitly. Quality requirements are typically stated in a disjunctive way, stating what should not take place rather than what should be achieved. In an example from the author's experience, a contract clause concerning reliability required no more than two service calls per 5000 operating hours in a medical instrument being developed. When stated as guidance for the engineers developing the instrument, the reliability clause really asks the instrument to operate without failure for at least 2500 hours at a time. It is left to the judgement of the engineer what failures might occur, what their effect on reliability might be, and how their incidence could be avoided.

From these two complementary kinds of design goals, performance and quality requirements, a useful lesson about the activity of engineering design can be drawn. In most cases, the client leaves a large portion of the decisions to the designing engineer. The author has even been involved in projects where the first contract paid the design organization to write the specifications which later became the basis for a second contract covering actual design and development.

The engineer is obliged to solve the immediate problem of the local details at hand while meeting the needs of the global solution as a whole. The engineer is obliged to coordinate information and the results of decisions with other engineers doing other parts of the design. The engineer is obliged to look for possible quality issues within the design, which is equivalent to inventing problems where none are obvious, and fixing them.

Engineers plan for the allocation of scarce resources. Ordinarily the resources are assumed to be those used as raw materials in engineering processes: water, energy, materials of construction, etc. It is important to recognize that engineering labor hours are themselves a scarce resource, and engineers who are busy making design decisions with scarce material resources are simultaneously trying to allocate scarce personnel resources. There are constraints, provided by the client, on how much time and money will be allowed to build a unit of the product being designed. There are also constraints, not as obvious to the outside observer but equally plain to the engineers involved in the project, on how the product development team spends its time deciding how to meet the client's needs in the context of the remaining technical and management requirements to be met. In essence, engineering design is the art of making correct decisions despite an overconstrained problem.

Although it is not entirely correct to do so, it is convenient to break the design process down into successive phases. A few of the best-known descriptions of the sequence in design from the literature will be mentioned in the next chapter. The sequence given here is the author's version, built upon not only the literature on the subject but also on experience with management of design projects.

In broad terms, the design process refers to conceiving, developing and detailing the solution to the posed problem. There are interesting aspects, worthy of research, in all three of these phases.

Conceiving the solution may also incorporate effort to further develop the statement of the problem. At the concept stage, the solution is described in enough detail to show that it will meet the principal functional need it is intended to address. Ordinarily this means the entire design is notional, although it is common that a portion of the design involving a critical technology will be developed in some detail in order to demonstrate feasibility.

Developing the solution is the process of refining the concept to the point it appears to be workable, achievable, and minimize risks, all under the costs defined in the contract. Developing the solution may involve developing several competing solutions, then choosing one. Developing the solution normally includes analysis work to verify the contract requirements are being met. At the development stage, there is only enough detail present to permit the analysis to take place. Normally the analysis of the design during the development phase is predictive in nature: the assessment is intended to lay out sizes, shapes, and fits based on expected loads and service life.

Detailing the solution refers to moving the description from a basis for analysis to a basis for fabrication. Some of the details present during development are no longer at issue, a lot of additional detail will be added. At present, a majority of design for manufacturability (DFM) emphasis is being applied to the detailing phase. There is substantial room for incorporating DFM into the development phase, illustrated by the exercise described in Chapter 4.

An old adage refers indirectly to the sequence of the design process: "if you understand the problem, you're most of the way to the solution". The expression implies that a complete problem definition is the first logical step in the product development process. The author disagrees with this notion. Problem definition is a long process, and one which cannot be completed before the design is begun. Part of the problem definition process cannot be accomplished without the context of the design being developed. A radically different solution approach would lead to a different focus on the problem definition. The adage should be reworded thus: "by the time you understand the problem, you're most of the way to the solution". Problem definition is difficult and can be expensive; and the problem to be solved sometimes will change over time.

A design problem from the author's professional experience will illustrate the point. Recently a device was built to collect oceanographic water samples; the author was a member of the team which carried out concept development, detailing, construction, testing in the laboratory, and testing at sea. The effect of a wave slapping against the exterior was a concern, but no data which seemed useful for this specific application was available. There were handbook values for the pressure exerted by a wave on the hull plating of a ship. An effort was made to instrument a spar in the ocean, subject to waves, but the spar was near shore, in breaking waves, and data obtained do not represent open-ocean, non-breaking waves. Only after the first tests of the original design was the wave pressure problem understood well enough. Redesign led to a solution to the problem. Testing at sea costs a minimum of $25,000 per day, and it was decided the risk of not succeeding at the first sea trials was worth the money to be saved. The cost of defining the problem is one of several reasons that problem definition was developed in parallel with the design itself. In this example, the device was built and being tested before the problem definition could be complete.

Much of the literature refers to design as an iterative or cyclic process. In this author's experience, some projects can take on a cyclic form, while others seem to be a linear process. The size of the cycle can also vary, with the iterations dominating the entire development or merely small eddies, found in a subset of the decisions which are made.

There appear to be at least two different reasons for the appearance of cycles in a design project. One reason is that some decisions are made iteratively, with deliberate approximations requiring successive refinement.

Another reason is a decision may be found to be incorrect or to have been invalidated by changes in the nature of the problem, and must be revisited.

Iteration takes place when it is necessary to guess at a portion of the solution. It is quite common to face a circular situation in which a portion of the solution is needed to develop other portions, and the other portions in turn are needed before the initial portion can be clearly defined. The engineer uses some judgement and makes some form of estimate to initiate the process. Later, the circle of reasoning returns to the initiating value, and it can be refined based on the decisions made in the interim. But there are decisions which were dependant on the original decision, and when it is refined, it is necessary to refine the dependencies as well. These dependencies in turn circle back to another refinement of the initial value.

In mathematics, an iterative solution is only possible when the solution series converges on a single point. In design, a number of judgement calls can be made to limit the need for further iteration. A common judgement to stop further work is recognizing the initiating value has changed little enough. The point on which the solution has converged can be considered in terms of the physical meaning of the mathematical solution, allowing the engineer to discriminate between multiple possible solutions. Even when formal convergence isn't possible, engineers can use iterative solutions effectively when making design decisions.

It is also common that a decision is made, which later is shown to be inconsistent with the design. Such a decision is not necessarily an error by the engineer, although errors are certainly made in design, and usually detected and corrected. It is also possible the decision was made before a change in the requirements which made it no longer valid. Assumptions are sometimes necessary to advance the design despite some roadblock; later the roadblock is cleared and some of the assumptions must then be revised.

Design problem-solving is characterized by the following features:

- The solution takes the form of a plan showing how to fabricate and operate the solution to a problem.
- Each decision must advance the development of the plan, either by adding to the global structure or by adding local detail.
- There are a large number of requirements which must be met.
- Some important requirements are not stated explicitly, in particular the requirement that unexpected failures be avoided.

It might seem surprising that engineers are capable of functioning, given the number and nature of the constraints in a design assignment. The reasoning processes which permit success in this complex environment is the subject of the next section.

2.2. Reasoning Methods in Design

If simplifying the problem is the fundamental approach to design problem-solving, then inexact methods of reasoning form the principal approach to those simplified problems. There are many methods for inexact reasoning in engineering design. These include extrapolating from experience into an unproven area, relying on a model to determine properties of the design, using approximations which simplify calculations, using plausible inference, and reasoning by analogy. These methods are recommended by writers about methods for solving engineering problems (Alger and Hays, 1964; Middendorf, 1986; Bailey, 1978), from which two critical properties of inexact reasoning methods emerge. First, the reasoning method is clear and simple in its application to specific problems. Second, the process of learning to use the reasoning method develops an understanding of not only the technique itself, but also the limitations inherent in it.

Reasoning methods which provide specific methods for specific, limited situations, and guidance on those limits, have been recognized and named. They are called heuristics (Koen, 1985). The author agrees with Koen that all of engineering thinking is characterized by the use of heuristics. Inexact reasoning is synonymous with the use of heuristics.

An obvious example of inexact reasoning in engineering is the number of decimal places used to define the number π (Pi). The author memorized 3.14159 years ago, but knows of others who have memorized 3.1416. There are situations where 3.1 is sufficient, and this value is fairly common when using a slide rule or calculator without the value π already supplied on it. When the author was making a tube out of flat, heat-sealable material, a value of 3 was sufficient for making a quick, mental-calculation estimate of how much flat material would be required to form a tube of the desired diameter. In the case of the tube, the time required to go find a calculator was not worth the improved accuracy.

To better understand the two properties of inexact reasoning, consider the theory of elasticity, such as that given by Timoshenko and Goodier (1970). A series of underlying assumptions are made about the general behavior of elastic materials, such as isotropy and a linear relationship between stress and strain. On the basis of those assumptions, specific procedures and methods are developed to deal with specific classes of problems. To apply the theory based on these assumptions to a particular problem, it is necessary to reconsider how well the assumptions apply to the particular case being solved. Timoshenko reexamines those assumptions in more detail in a section late in the book, dealing with more complex problems where the assumptions are no longer obviously valid.

To solve specific problems more conveniently, additional assumptions can be made. Examples of simplifying assumptions are the assumptions of plane strain, at the center of a thick section, or plane stress, at the surface of a section, or throughout a thin section. From an absolute viewpoint, these assumptions are not valid, but they do represent the two situations with sufficient accuracy to allow reasoning to continue. Plane strain or stress assumptions make calculations easier by reducing the numbers of parameters and eliminating several terms containing cross-products.

Not only are there specific methods of solution, and underlying assumptions imposing limits on validity, the two are intermixed and cannot be entirely separated. In the derivation of a method, the assumptions are stated at the point they are needed for the method to be further developed. An assumption is not invoked until it is necessary to make the method easier (or possible). Similarly, in use, the assumptions are invoked and checked for validity implicitly; only rarely is there an explicit rule given for checking that a condition is being met. Watching one familiar with the methodology at work, one might form the impression that the methods were all clear and unambiguous. Listening to a lecture on the subject, one can easily overlook the assumptions by focusing on the methods and their applications. However, the limits on validity are always present and must be checked, at least qualitatively. (The most common qualitative check on validity is reasoning by analogy to similar problems for which the method was valid, such as solving a problem while following a worked example in a textbook.)

An example from the author's experience will illustrate application of several inexact reasoning methods used in conjunction. The author was developing the design of a selector valve which would direct one of several possible sample ports to the suction inlet of a pump. It was necessary to determine a method of sealing the moving rotor, which did the selecting, to the valve body, which contained the sample ports. Of particular interest was the energy required to actuate the rotor to index from one port to the next, and the amount of leakage past each seal.

First, a geometric model was build, using a CAD system. This simplified the reasoning necessary to determine appropriate dimensions of the unit and verified that the concept was feasible. The valve was simplified for modelling as sketches of several flat-plane sections through the cylindrical valve. Second, a physical model was built, identical in certain key dimensions to the intended valve, but having fewer ports on both the valve body and the rotor. Seal leakage was inferred by monitoring the water level inside the valve body, at a location where only leakage could have an effect. The effort required to index the rotor was measured with a torque wrench. A series of measurements were made to determine the effect of seal interference on leakage and torque.

Third, the data collected was graphed, seal leakage and torque plotted as a function of seal interference. A graph is a model based on functional behavior instead of physical appearance. By inspection of the graph, a design was selected which compromised between limiting leakage and realistic actuating torque. It might have been possible to formally optimize the design, but such an activity was not considered an economical use of design resources.

The inexact reasoning methods which the author used to deal with the selector valve problem included reasoning by analogy, extensive use of several types of models, and plausible inference. Each of the models used was an analogue of the full-sized valve, but the type of model was selected at each point in the reasoning process to obtain the particular results needed. Plausible inference, that is, using an unproven explanation which fits the facts, was used to determine where the optimum seal design was located. The physical model and the geometric representations all were analogs of the actual valve to be built. The valve has been built and tested and is now in service.

Inexact reasoning is a necessary aspect of design because, in most instances, exact reasoning is impossible. An exact solution method which fits the problem to be solved must be available. The solution method must have a form appropriate to the predictive needs of the design situation. The necessary data about properties unique to the design must be available. The necessary resources - time and personnel - must be available to carry out whatever calculations are required, and to evaluate the meaning of the evidence those calculations produce. Only rarely do all these conditions come together.

The use of models is such a pervasive method of inexact reasoning that it deserves particular attention. All engineering design problems are dealt with through models. This is especially true for evaluating a design to predict its performance. A model is an analog of the full-scale system being designed. There are several important classes of modelling methods. These include scaled-down physical models, either geometric (as in a scaled ship model in the towing tank) or by another system parameter (as in a chemical pilot plant, scaled down in production capacity); a numerical simulation, such as finite element or finite difference methods; a statistical representation, reducing a number of separate experiences to a collective distribution; an analytical representation, reducing the problem to a group of equations; or a graphical representation as in a sketch, drawing, or nomogram. The general concept of modelling represents a very powerful method of inexact reasoning.

What is critical in all modeling methods is that valid application of a model depends on the judgement of the modeler. More than one approach is usually feasible, and which is chosen depends on the purposes of the user.

Middendorf (1986) defines a model as "any representation of the proposed system or device that contains enough information to be useful in making design decisions." He goes on to note that early models are likely to appear crude and incomplete, but are refined as the situation is clarified. Economy plays an important role in the use of models, so Middendorf’s definition might be modified to indicate a model contains just enough information to form the basis of a decision. Exactly what constitutes just enough information in the model will change as the design progresses, and is changed by the engineer as the context changes. A change in the way a problem is to be represented may be reflected in the need to change the model used.

There is a connection between economy of use and how well the model reflects reality, as judged by the immediate decision-making needs of the engineer. At each point where modeling is necessary, some set of issues must be resolved before the design can proceed. Even if the model is incomplete, or relies on an "incorrect" representation of an issue, it may still serve the need of providing evidence required to resolve the issue at hand. To achieve economy for a given decision, one chooses a model which adequately mirrors reality, but without excessive detail in the representation. To achieve economy in general, one requires an adequate selection of candidate models to choose from, each appropriate to different situations, and an understanding of how each can be used.

For this reason, there is no such thing as the "right" answer to a problem. In many cases the modeling conditions preclude the existence of a single answer. In still-water model testing for ships to determine the power required for propulsion, for instance, two dimensionless numbers are important in scaling the model's fluid dynamics to the behavior of a full-sized ship. The Reynolds number, dealing with water friction over the wetted surface of the hull, is related to the wetted area of the ship and the model. The Froude number, dealing with the energy dissipated as waves generated by the ship's form, is related to the length along the waterline. These two scaling parameters, in general, cannot be met at once (Comstock, 1967), thus it is necessary to accept an approximation in the use of the model. In tests involving ship motions in waves, details of the shape of the hull grow in importance, making scaling more difficult still (Bhattacharyya, 1978). The engineer must judge the weight to be given to each parameter involved, and must judge how well the results can be scaled to predict full-scale performance.

The value of a result at some intermediate point during development of the design depends only on the purposes of the engineer developing the concept. The logical progression of a design can be compared to a mathematical proof in that there are, potentially, multiple paths from the initial conditions to the conclusion. The progress of a design development, just as in the progress of a proof, indicates the engineer's style in approaching problems and will influence the cost of obtaining the conclusion, but has little bearing on the validity of the final outcome. In mathematics, a short proof is admired as elegant. In engineering, rapid progress to a valid design is equally elegant and deserves equivalent admiration. But neither mathematics nor engineering include consideration of the methods which could be used to achieve these economical solutions in the mainstream of their research.

The decision about which part of a system does which tasks is known in decision theory as "allocation of function", the human actor being part of the decision-making system. Price (1985) is convinced there is no procedure for allocation of function, and suspects that for most systems there is no fixed allocation which is valid for all situations the system must be expected to handle. The question of selection of a model - or selection of a design tool - is an allocation of function problem. This is best handled by the engineer, in the context of the problem to be solved, the immediate decision to be dealt with, and the desired economy in that decision.

The use of inexact reasoning, or heuristics, is the basis for much of the work reported in this thesis. Inexact reasoning is the basis for a vast majority of the problem-solving accomplished by engineers.

2.3. Design Quality

Improving the quality of manufactured goods is a topic which is currently receiving extensive attention. It is believed that improving and maintaining high product quality is the key to American industries remaining competitive in the international marketplace. As an extension of concern for quality of individual products, attention is also being given to the idea of designing for product quality.

Improving product quality is currently a major preoccupation in American industry. Traditionally, Quality Control has been an inspection program after fabrication was complete. Only recently have there been efforts to build quality into the product. Quality can't be inspected into the product after fabrication is finished (Box and Biggaard, 1987). This comment is corroborated by the observation (Crosby, 1979) that in some factories a large portion of total production - possibly as high as 30% or 40% - is generated by rebuilding products which had originally been rejected by Quality Control. Rebuilding rejects is a costly way to make the product, and getting the product right on the first time through production is a major emphasis in industrial management. There are, of course, some products with a combination of difficult processes and high standards - such a directionally-solidified superalloy turbine blades - that high rejection rates are common. High reject rates or high costs of manufacture are not themselves to be avoided. What the engineer does wish to avoid is the unpleasant surprise of unexpected high costs in manufacture, and unexpected quality problems. If the part truly is going to be hard to make, then that difficulty should be a deliberate design decision and not a surprise during initial production.

As responsibility for quality has moved back upstream, the means to obtain quality has undergone a major change in emphasis. Rather than inspection, there are now approaches based on encouraging quality workmanship among employees, such as "zero-defects" (Crosby, 1979). In terms of engineering design, most of the emphasis has been on making the designed parts easier to inspect, although design for Quality Control is not the same as design for quality products. The discussion of designing to achieve product quality has only begun (Box and Biggaard, 1987).

A logical extension of an interest in improving product quality is to consider the design a product, and to examine quality of the design. Design quality as a concept would include all factors affecting product quality, but can also include aspects not normally associated with the product itself.

The concepts Taguchi (1979) has developed for product quality can be applied to design quality. In particular the definition, which follows, of loss of quality in terms of costs of non-performance is appropriate. Quality is difficult to measure or improve directly, but loss of quality is measurable and the measures provide guidance on how to reduce losses. Loss of quality in a product can be defined in terms of costs due to the unit failing to serve its intended purpose.

Every product has intrinsic value, associated with what and how it produces something for its owner. A factory tool has value not only because it was a costly capital investment, but also because of the value of the products it helps to create. Should the tool fail to operate, production is slowed or stopped until a correction is made, and the cost of lost production begins to accrue. Even if changes can be made to continue production, there are associated losses to the company. There are costs to the client who loses access to the product, and the suppliers who provide the materials the production line uses to make the product. These costs, loss of value from the machine, represent a way of measuring loss of quality in the machine. Similarly, a car which fails to start one morning generates costs for its owner, and possibly for others as well, such as the car owner's employer. A machine with ideal quality never loses value for anyone affected by its operation. A machine with poor quality, conversely, generates losses of value. Regardless of who suffers the cost, loss of value adds to the reduction of quality in that product.

Product quality can be measured in terms of the costs (to anyone) of a failure of the product to perform its function. In the usual sense, this simple connection of ideas provides some specific measures of product quality. But the philosophy behind Taguchi's definition of quality can be used as a broad basis for thinking about the design itself.

The design is the product of an engineering process. As a product, it has users - most notably, the product manufacturer and the person buying the product for use - and those users can experience costs when the design breaks down. Obvious breakdowns are those dealing directly with product quality: designed-in failures, poor access to maintenance items.

Less obvious breakdowns of the design can affect the cost of production, including details which have been left ambiguous, making it necessary to stop and figure out what will be needed. Another example is in dimensioning a drawing; the dimensions an engineer needs to carry out an evaluation or to coordinate one part with another are not entirely compatible with the dimensions the machinist needs in order to make the component, and the drawing's format can increase the likelihood of mistakes.

Taguchi (1979) has developed methods to plan for quality production, which he calls off-line quality control, which involves quantifying quality aspects of interest through figures of merit, then conducting statistical experiments to determine where the critical points in production lie. It is natural to extend planning for quality to the concept of design quality as well. In the fourth chapter, discussing design for manufacturability, the author will demonstrate figure-of-merit comparisons for use during material selection. But use of purely quantitative comparisons is limiting. In the fifth chapter, discussing design for product reliability, the author will describe predominantly qualitative methods of dealing with failure prevention. The balance between quantitative and qualitative reasoning for a problem is driven mainly by the context of the problem itself.

Statistical process control methods are based on getting feedback from the process to be controlled. In design, such feedback is often incomplete, or slow to arrive, and may be simply unavailable. Throughout this research the author has emphasized predictive, "feed-forward" approaches, relying on knowledge which helps the engineer anticipate issues which are important. The expression "proactive management" enjoys popularity currently to describe a management style which anticipates and acts to prevent problems before they emerge, rather than simply reacting to problems after they are evident. By the same usage, proactive engineering is anticipating problems and correcting them during the design process, rather than when they become evident during manufacturing or in service.

Design quality deals primarily with interactions outside the design process - manufacturing, service experience, maintenance, retirement of the unit. For that reason, design quality can only be accurately measured years after the product development engineer has finished work, when the evidence required becomes available. Since rapid feedback for a closed-loop system is not available, design is of necessity an open-loop situation. The closest thing to feedback is generated by experience with other, similar designs, and through the use of models. The engineer exercises control over the outcome by a conceptual, predictive understanding of the events to be controlled, not by tuning based on feedback.

Design quality is a way of describing and dealing with virtually every improvement to the design process which has been made, and which are forecast for the future. While the knowledge described in the examples of the fourth and fifth chapters has some interest, it deals with specific problems, and the more important product of this research is demonstrating general approaches to design quality.

2.4. Engineering Productivity

Managers in industry say that improvements in product quality are improvements in engineering productivity (Webster, 1988; Decker, 1983). That happens to be a problem of current interest, just as manufacturability is a focus of attention. Product quality and manufacturability are both clear examples of the need to raise the quality of our designs, and to make design decisions more effective, not just more efficient. When implementing improvement to the ways current problems are handled, it is important to make sure prior solutions are not adversely affected, and improvements to deal with subsequent problems can be accommodated within the same general framework.

In terms of economy, or engineering productivity, two different criteria are available:

- Efficiency measure: are designs being generated with less use of resources, without reduction in the quality of the design?
- Effectiveness measure: are higher-quality designs being generated?

The first concept is generally used as the only method of predicting improvements in design productivity. The second, introduced by the author in Chapter 2, is another dimension of productivity which is important.

It is necessary to improve both efficiency and effectiveness of design decision-making in order to accomplish a substantial change in engineering productivity. In the chapters to follow, two very different methods for improving the quality of designs will be considered.

Engineering productivity improvements which deal with more effective decisions deal with improving the quality of the designs produced. Productivity of engineers is a primary justification for the investment in CAD equipment. In fact, much of the advertising for CAD warns that further investment in CAD equipment is the only path to improving productivity.

The usual measure of the value of an investment in any capital equipment is the return on investment. The simplest version is marginal return, or the increase in income generated by the investment. Other investments are evaluated by examining tradeoffs, which account not only for the cost of equipment and the income generated directly, but also indirect costs and savings caused by the change. For industrial automation equipment, cost savings can be explained in terms of fewer people doing the same job, or the same people doing more work. Productivity is usually taken as meaning the rate of production. By investing in automation equipment in the plant, management expects to have lower costs per unit produced, or a gain in productivity. Another term for such a cost reduction is efficiency; "efficiency experts" have had a steady business of inspecting factories and recommending changes.

But efficiency is not the only aspect of improving productivity in engineering decision-making. Improving effectiveness of design decisions is even more important because design quality is affected. The largest gain needed is to improve the favorable outcomes of decisions, that is, to make decisions more effective. An effective design decision is one which does not require correction or redesign later in the process. Improvements in the American automobile industry have emphasized product quality through more effective design decisions. Problems with designs which are not manufacturable, for instance, reflect ineffective design decisions, irrespective of the efficiency with which those decisions were made. The effectiveness of a decision refers to how well the results meet the design requirements. Efficiency in a decision refers only to how resources are used when making it, and therefore has no relationship to whether the decision was effective.

A frequent theme for improving productivity in some organizations is the need to "work smarter, not just work harder". In that idea is the same distinction between effective and efficient decision-making. One could describe the research being discussed in this thesis as exploring means to allow engineers to work smarter. Most developments in design tools, however, seem to be targeted simply on working faster. Rapid drawing time is a main selling point of graphics CAD systems (Chorafas, 1987). Rapid analyses, or more analyses or more detailed analyses in the same time span are cited as principal features of finite element packages (Berry, 1988; Braham, 1988). Rapid access to a larger information store is the principal purpose of a wide range of databasing tools and methods for engineering (Winston, 1984; Yeomans, Choundry, and ten Hagen, 1985). Working faster or working harder, efficiency, is clearly the main thrust of efforts to improve engineering productivity. A more detailed discussion of computer aids to engineering will be the subject of the seventh chapter.

The author's initial experience with CAD is the same as with power tools for woodworking: they enabled him to make mistakes at a faster rate. The presence of a power tool alone does not necessarily improve the design, although progress can usually be made at a faster rate. In the author's experience with both computer support to drafting and computer support to word processing, most of the time savings which might be realized from working faster is consumed in a series of corrected versions. It is necessary to make the original decision more effective, so the first version is correct, in order to achieve a net improvement in productivity.

It is instructive to examine how a design tool actually contributed to one company's productivity. The firm makes molded plastic parts, primarily for the automotive industry, earning about fifty million dollars annually in gross sales. A graphics CAD system was introduced a few years ago, which the author discussed with managers, engineers, and draftsmen. Their initial motivation was to simplify drafting, and to improve the ability to revise drawings to adapt an existing part to new requirements. Drafting efficiency was improved through the use of the CAD system, but that was not the primary improvement, and their gains in that area were not as large as they had expected.

The system the company purchased was capable of handling three-dimensional drawings, and the drafting department developed approaches to take advantage of this feature. Some of their parts are molded in one configuration, then folded on assembly, such as clips to hold several hoses or cables together. On paper, the drawing of the folded part rarely was more than a sketch, and great effort was necessary for the draftsman to correctly render the unfolded version for the molding process. The draftsmen discovered they could, on the CAD system, draw a part in its assembled form, then unfold the drawing on-screen to obtain the mold pattern. This simple change has greatly reduced the complexity of drawing folded parts, and eliminated many errors. Managers at the firm consider the ability to draw and manipulate three-dimensional objects to be a productivity gain, because of the increase in quality, even though no appreciable cost reduction can be associated with it. They anticipate this quality improvement will lead to cost reductions in the future, primarily through a decrease in time and costs of prototype runs.

The CAD system contained the basis for communicating with CAM systems which the vendor was not yet listing as a working feature, as it was still under development. Within one year after introduction of the CAD system, the company was able to produce, on numerically-controlled machines, prototypes which had in the past been made by hand in the toolroom. The CAD system could redefine the drawing to show the space around the part instead of the part itself, allowing injection-molding dies to be specified directly from the CAD system as CAM instructions to their subcontractors instead of as drawings. This change moved control over the process of producing dies closer to the engineer responsible for the part, but has not substantially reduced the cost of a die.

The company is pleased with their purchase. The improvements they have achieved are significant gains in a highly competitive industry. But the gains promised by the salesmen were not realized. In particular, drafting time was not reduced significantly, even in the case of adapting an existing drawing to a new purpose. The productivity gains anticipated by the company's management were not met either, no large cost reductions have been obtained. Of the two major improvements described above, the first was discovery and exploitation of a minor feature of the system, and the second was adapting and using a feature the system's vendor did not consider complete and was not supporting.

The major gains obtained through the use of this system dealt with more effective engineering, instead of more efficient. In other words, designs were produced no faster, but better designs were produced.

This company's experience illustrates two points. First, the design tool developer is not in a position to anticipate how the system will create improvements in productivity. Only users can find features of the system which are valuable and adapt them to serve their needs. This point will reappear as a major emphasis of the seventh chapter.

Second, engineering productivity is not necessarily measurable by the usual cost-reduction formulas. Downing (1989) makes the same point. This outcome is related to the way costs and benefits of the tool's use are measured by management, not to the usefulness of the tools introduced. Although generally unable to show the improvement in terms of tangible costs and benefits, engineering managers interviewed by the author and by Comer (1987) feel that installing CAD has in fact improved productivity. A primary reason is that these engineers do not measure productivity through the limited means available to accountants.

The author has visited several organizations which had assumed productivity would mean design could be handled by a smaller staff. But after a tool for productivity, such as a CAD system, is installed, there is no less activity in the engineering department. In some cases the engineering staff has expanded since the CAD system was installed.

The typical CAD installation is oriented toward drawing, maintaining, and manipulating pictures of components and assemblies, and putting that drawing into the right data structure to communicate with other equipment. Such a system can reduce the time required to communicated complicated ideas which have, in the past, required large volumes of paper. It is clear there are many advantages to working with a computer-based drawing system. However, one must not assume that the efficiency gains offered by CAD is the only available improvement to engineering productivity.

One additional aspect of engineering productivity requires consideration: the cost of correcting errors. Efficiency of decision-making can be evaluated in terms of costs by accounting for the labor and resources which go into the decision. Effective decision-making can also be measured, using a definition much like the definition of quality introduced above. An effective decision is one which does not lead to corrections requiring additional, unplanned costs in the design. An effective decision is one which contributes to a quality design; that is, a design without defects.

From the data in Figure 2, the author concludes that the earliest, global design decisions are critical. The early decisions shape the whole project. Because of the high potential cost of correcting errors in an early decision (Fabrycky, 1987), it is vital that the decision be effective. In other words, preventing errors in the earliest decisions is a principal goal for engineering designers.

It would seem logical to expect the engineer's employer to make a substantial investment in support of effective design decisions during

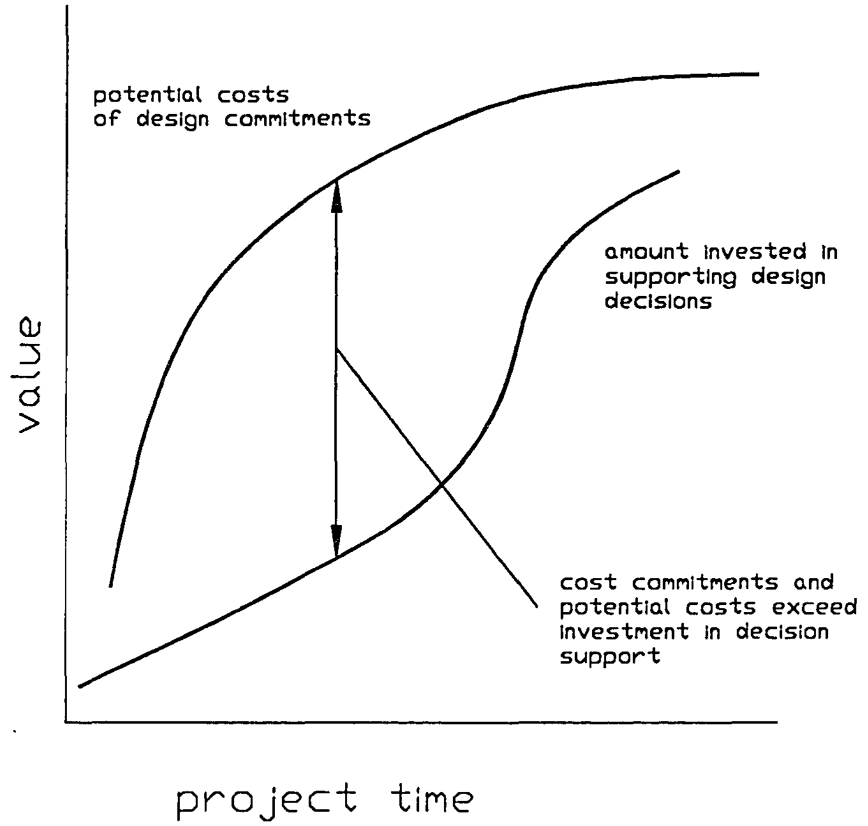


Figure 2

Decision Costs and Decision Support

(after Fabrycky, 1987, and Aries, 1987)

early, global development of a product. Figure 2 shows such early decisions receive the least investment in support of those decisions (Aries, 1987). The late decisions, which are localized, have considerably smaller potential cost, yet this is the area where large investments are made to support decisions. In a large project a portion of the design budget may even be expended in making calculations after the design is already under construction, to verify the design. But at this late point corrections (if feasible at all) would be very expensive.

The data showing investment in decision support in Figure 2 is taken from a company which sells computer-based aids to design. It may be assumed this data reflects the investment in tools for decision support, such as computer aids. The author has concluded above that even the knowledge supplied to engineers is not well-suited to design, and therefore even the investment in design decision-making support associated with ordinary research funding emphasizes late, details-oriented decisions.

There is a clear economic incentive to develop supports for early decisions. Effective decisions save money by preventing the costs of error correction later in the project. Effective decisions improve competitiveness by reducing time to market, cost of development, and by improving the quality of the product through the quality of the design. The costs of error correction are often a surprise, since the error is not a planned part of the product development process. Effective decision-making, and an investment in support of improving effective decision-making, makes projects more manageable by reducing unintended errors which lead to unplanned costs of correction.

2.5. Summary

In this chapter, the author has presented some thoughts on design. Design is the process of developing a description of the hardware to solve a problem. Engineers are normally required to develop their own problem definition based on the incomplete definition supplied by the client.

Engineers must meet design goals which may be explicit or implicit. The author has called the explicit goals performance requirements because they deal with how the product will perform or operate. Engineers must also search for, recognize and meet implicit goals, of which only a very limited number will be supplied by the client. These implicit goals have been called quality requirements because they deal ultimately with the quality of the product, or of the design.

Engineers plan for the allocation of scarce resources. During design, the resources allocated are not only materials and energy in the design itself, but also personnel resources to carry out the design task. Management of a design project requires attention to both technical progress and technical talent. The two are not separable.

Engineers solve problems through the use of inexact reasoning. Many forms of models are used to reason by analogy about the design. Two features of reasoning methods are important. First, the method is selected because it provides economy; that is, the answers needed are obtained through a minimal investment of time and effort. Second, the method is selected through an understanding of how closely it mirrors reality. Selection requires a compromise: a more refined, more detailed model will generally require more effort to build, operate, and understand well enough to obtain an answer.

Design quality is a way of thinking about improving individual products - that is, product quality - and improving individual designs, and improving the design process. A design which is of low quality has higher costs to implement, just as a product of poor quality imposes unwanted costs on its producer and it user. Quality of a design is a convenient measure of whether a suggested change to the design process it worthwhile.

The concept of design quality can be used to reason about engineering productivity. Normally, productivity invokes an image of efficiency, that is, of working faster. For engineering, making decisions at a faster rate runs the risk of making mistakes at a faster rate as well. An increase in the rate at which a design is produced must not lower the quality of the design. In order to really advance engineering productivity it will also be necessary to improve the effectiveness with which decisions are made, in pace with development of improvements in efficiency.

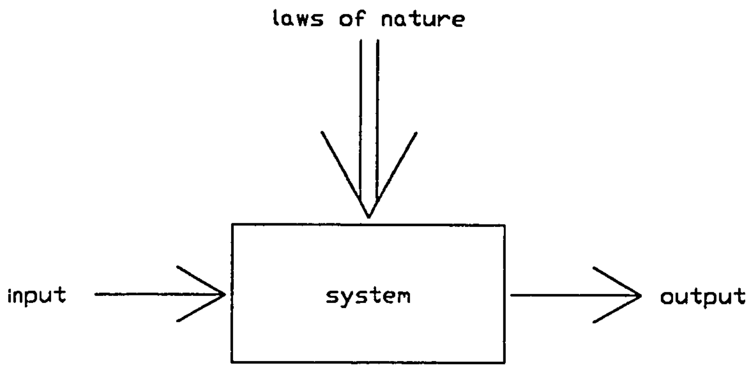
Chapter 3
Design Theory

This chapter contains a review of the literature of design theory. Design theory is necessarily a wide-ranging concept. The discussion to follow will examine the ideas and contributions to be found in a wide variety of disciplines which deal with design and with solving problems.

It is instructive to begin by examining the nature of design thinking. Figure 3 illustrates the general modes of thought when solving technical problems. The diagram shows a system, influenced by the laws of nature and by inputs, and producing outputs. A typical homework assignment in an engineering class would state a system and its inputs, imply the laws of nature, and ask the students to determine the outputs. Dixon (1963) calls this type of problem analysis; it is also called a direct problem.

When the problem is to find a system which produces a given output from a given input, Dixon (1963) refers to the problem as one of synthesis; it is also commonly called an inverse problem. This use of the words analysis and synthesis is not entirely in keeping with Asimow's (1962) definitions, but it is consistent with the most common usage noted in the first chapter.

The expression "inverse problem" implies that the thinking required to accomplish design is to work backward through the method for a direct problem. This interpretation assumes that since direct problems are usually expressed as equations or algorithms, then design can be expressed as resolving the



GIVEN:	TO FIND:	IS CALLED:
input, system, output	laws of nature	a research problem science
input, system, laws of nature	output	a direct problem analysis
output, input, laws of nature	system	an inverse problem synthesis

Figure 3

A Diagram of Engineering Processes

(after Dixon, 1963)

equation for the unknown, or working through the algorithm backwards. However, it is important to note that many equations cannot be solved in closed form for a variable which is embedded in the direct expression. Most algorithms make no sense whatever if read from ending to beginning. When rearranging an equation in order to accomplish design, it is common that there are several unknowns for which assumptions must be made. Still more common is the fact that much of the design is not solvable via analytic expressions at all, and the inversion is actually a very complex twisting of direct-problem concepts which the engineer must carry out mentally. For instance, the graphical method of analyzing hydrodynamic bearings found in a machine design textbook (Juvinall, 1983) cannot be operated in reverse to develop a design from a desired outcome. Instead, an iterative approach is necessary, which must be seeded by educated guesses. The iterations required may fail to converge on a viable solution if the assumptions made along the way are not reasonable.

One must not be fooled into thinking that design is simply a mathematical mapping from direct problems into another solution space. Design is quite different from direct problems in the nature of the problems as well as the nature of the solutions.

There is a prejudice among those who produce new engineering knowledge toward the direct problem, revealed in the very use of the term. Most knowledge for engineering is recorded in forms amenable to direct problems - that is, to problems involving evaluating the behavior of a known system. Knowledge which can be used for direct problems is not necessarily of value for inverse problems, where the system must be created to meet requirements.

There is a need, therefore, for predictive design knowledge, to be extracted from that knowledge which is already available for analysis.

3.1. Literature and History of Design Theory

Design has always relied heavily on the individual engineer's experience. Recently there have been changes in the way engineering is undertaken. These changes include greater limits on the acquisition of experience, greater emphasis on use of complex methods for more refined predictions, greater requirements to effectively incorporate new technologies, and an emphasis on engineering results having a sound, clearly definable basis. These provided the driving force for the development and study of design theory.

To a certain extent, the history of design theory is a collective history of both the changes in the ways engineers approached problems, and the economic and social forces which brought about those changes. One can divide the development of design theory, all of which has developed in the last fifty to sixty years, into five main periods. Each of these periods provides some useful insight into the nature of design. Although these periods overlap and in most cases have not closed, in rough chronological order they are:

- the earliest era, dealing primarily with experience as the basis for good design;
- the development of structured approaches to design, concurrent with the advance of the concept of engineering science;
- approaches to automation of design, based on advances in computer technologies;
- application of decision theory developed in economics, psychology, and philosophy to describing engineering decisions; and

- return to a broader concept of design, in which the theory is more like a philosophy and less like a procedure.

The last two aspects have developed more or less at the same time, in the past few years. In this section, the author will review and comment on the literature, grouped generally by the categories given above.

Traditionally, design has been based on the individual experience of the engineer. This tradition stems from the period, perhaps a hundred or more years ago, when design, engineering, and production were all blended together in technologies dominated by craftsmen. Prior to the advent of mass production in the factory, most engineering was carried out by masters who were fully involved in their projects, from concept through fabrication and operation, and who had considerable experience with similar projects. James Watt and Thomas Newcomen built their own steam engines; George Washington Roebling took equal interest in making the wire for the suspension cables and building the stone piers of the Brooklyn Bridge. With the development of mechanization and mass production approaches, engineering gradually became specialized and moved to locations outside the production facilities.

Experience remains, however, a dominant force in the development of engineers. In the past, experience was the most important component of an engineer's professional preparation. Even within a single project, a significant amount of prototyping and testing provided working experience with the system, which in turn was the basis for further improvement to the design. In addition, engineers tended to stay with a single company, or at least with a particular product type, for their entire careers.

Writers on design during this early period concentrated on experience; in fact, experience played such a large role that most early books on design take the form of memoirs. Glegg (1969) made a limited number of comments about the design process, and relied heavily on case studies to convey a sense of his experience in design. Teague's (1940) book is as much a memoir of his lifetime in industrial design as it is guidance for designers in other fields who wish to emulate his success. Neither the anecdotal approach nor his artistic, non-technical background interfere with his book's valuable contribution to an understanding of how design decisions are made. Teague's most significant contribution is a discussion of the old adage, "form follows function", which he extended to incorporate the concept of fitness - fitness to materials of construction and to intended users as well as fitness for purpose.

Shifts in the industrial economy after World War II had strong influences on engineers and their working lives. The first was the physical mobility brought about by career mobility. Mobility brought about a flux which is still seen today, as corporate decisions move a company into or out of a market, and subsidiaries are bought and sold. Increasing mobility in the American culture, driven by and driving the tendency of large companies to reallocate their human resources to meet changing needs, brought about instability in the process of accruing experience by providing interruptions and shifts in focus. Second, growing organizations had growing management needs, and many good engineers became managers at about the time they had enough experience to be really effective at design.

A third influence changed American education as well as engineering, the post-war (and especially post-Sputnik) emphasis on science as the foundation of knowledge. Engineering became dominated by engineering science, with the reasoning that engineering is primarily applied science (Alic, 1988). By the

1970's, engineering science was the dominant factor in accreditation of engineering curricula, as it is today.

The "engineering science" approach led to the next stage in the development of design theory: attempts to find scientific structure in engineering methodology. Nadler's (1967) work is typical of early structuring approaches, attempting to find an analog to the scientific method in design methods. Jones (1963) recognized that most design decisions proceed based on judgment, and discussed approaches to making these judgments more precise and more systematic, in order to obtain more consistent success in design. In the same conference, however, Christopherson (1963) noted that "design is not on the whole an orderly process".

One of the key concepts to emerge from initial exploration of the structure of design is the morphology of the design process. Asimow (1962) provides the most widely quoted morphology, or product development time sequence, which incorporated feasibility study, preliminary concept development, detailed design, production planning, distribution, consumption, and retirement of the product. He pointed out that each project is unique, however, and that emphasis can change as the project proceeds.

Asimow lists three forms of thought in design: analysis, synthesis, and evaluation; which represent understanding the problem, developing proposed solutions, and checking the validity of those solutions. These concepts are often mis-stated as a sequence in time as well, but his intent was simply to indicate the modes of thought involved. These modes of thought are difficult to separate from one another, since all three are required throughout the product development process. French (1986), for instance, deals with a wide range of concept-generation problems through analytical means.

With the advance of engineering science as a dominant philosophy in engineering, the division of engineers into more specialized, more isolated compartments became evident. The new specialty of systems engineering grew as a response to the need to deal with problems that crossed disciplinary boundaries, especially evident in design of computer and communications systems. Hall (1962) typifies the early concepts of systems engineering, which emphasized applying Information Theory to provide for control of individual, specialized contributions to a design. Systems engineering was originally conceived as the discipline which would coordinate specialist inputs to obtain a complete system.

As a strong mathematical and scientific basis for engineering developed, it was observed that there appeared to be considerably less uncertainty in the design situation, because technological advances provided means to predict more effects, and in greater detail, than had previously been possible. This observation led to the push for "rational" design, such as that of Weck (1966), in which decisions would be based entirely on accurate predictions, without the need for such "superstitions" as factors of safety. A factor of safety might be called a factor of ignorance, since it represents a margin left to accommodate unknown or unanticipated events. One must appreciate that in a proposal for rational design, "rational" has a narrowly defined, rigorous meaning which is quite different from the commonsense "rational" which means simply the result appears to be based on sensible, sane reasoning. In general, while "rational" approaches to design have not been extensively implemented, in many codes and standards there have been changes to allow design with lower safety factors in the few cases where the situation can be defined quite clearly.

Purely rational problem-solving, based on solid, logical reasoning, characterizes a mature science. But engineering design is always working at the edge of what is known, because that is where innovation is accomplished. For that reason, engineering can never rely exclusively on rational design.

The developing science of psychology has given a scientific basis for exploring some of the non-analytic aspects of design, especially creativity. Engineers have had a fairly steady supply of books on creativity enhancement, among them Gordon (1961) and Bailey (1978), which give a variety of methods for improved creativity to assist in developing potential solutions to a problem.

Operations Research is a field of engineering science which developed rapidly and prominently after the Second World War. Two main ideas were adapted into engineering design, the concept of formally optimizing design solutions and the ideas related to the theory of searches. Optimal design and optimization of both the product and the production process have become important to preserve economies in the large industrial corporations, and even in those cases where optimal solutions are not obtainable, a goal is to conduct an optimal search for solutions (Nevill and Crowe, 1974). Meredith, *et. al.* (1985) rely primarily on graph theory and optimization methods to deal with the problem of designing high-rise buildings and other large structures. Under the influence of Operations Research terminology, engineering design became divided into the search for solutions and the evaluation of those solutions, frequently referred to as synthesis and analysis. Although contrary to the meanings assigned by Asimow (1962), this usage reflects the meanings commonly attached to the words synthesis and analysis.

A book advocating the scientific basis for engineering is that of Middendorf (1986). He gives two main paradigms for problem-solving in engineering: design by synthesis (closed-form solution of a complete analytical model) and design by repeated analysis (iterative solution in a numerical model). Despite his strong inclination to a rigorous scientific basis in design, Middendorf acknowledges the continuing importance of experience, commenting that a distinguishing mark of a skillful engineer is the ability to develop designs which are close to optimal prior to evaluation.

Highly detailed, analytical assessments of the design situation will, as often as not, rule out all possible options. Engineering has a long history of cases, such as the Wright brothers' airplane, in which the design performed even though it was possible to "prove" analytically that no heavier-than-air craft could fly.

Some fields have had poor success with the scientific, rational approach to decision-making, reflected most notably in the literature of architecture and urban planning. These fields have expressed disappointment with rationality and have returned to more general, less structured (and less restrictive) approaches to design, placing more emphasis on the individual decision-maker understanding the problem in a way that provides for a creative solution. An excellent example of this change in attitude is given in Alexander (1984). Rittell and Webber (1973) deal with many of the reasons the scientific approach does not succeed in architecture.

The problems identified in the architecture and planning literature are equally present in mechanical design. Two main problems are largely neglected by those theorists who advocate highly scientific bases for design. First, design does not take place in a static world. Compiling the data required to define the problem fully takes time. As the problem and its solution are developed, the world continues to change. Some designs have become obsolete before prototype testing was complete because the need has changed, most notably military strategic weapons systems since the end of the Cold War. Second, the engineer is not only required to allocate the resources required to build and operate the product, but is also expected to economize in the allocation of resources during the design process. Compiling the data required to define the problem fully is a resource-intensive activity. The cost of data collection and the cost of decision-making are largely neglected by present research in design, one of the reasons that results do not match with predictions.

Computers have grown from laboratory curiosities to powerful tools in just a few decades. However, reliance on a computer as a tool in engineering influences the ways the engineer thinks and works. In particular, applications which use a computer must create a structure for the task, allowing it to be programmed, and that structure must be planned in advance. For instance, one must create the format for a database before entering the data, although in the author's experience the structure appropriate to the database is only apparent after at least some of the data has been recorded. In general, the structure required for computer use involves breaking a task down into isolatable subtasks.

Computer-oriented approaches led to an expansion of design theory into two areas: numerical models as a basis for problem-solving, and using the computer to record and maintain design graphics and data. With the development of artificial intelligence methods which could potentially be applied to engineering with reasonable effort, emphasis shifted to automation of the design process, in which the computer contributes to actually generating the design (Winston, 1984).

Furman (1970) edited an early volume dealing with how computers can and should be used in engineering. The main theme of the volume is that, by concentrating on evaluation in design, the design tool developer may be avoiding the real issues of design. Many of the preliminary concerns expressed by the articles in Furman's book have matured into genuine problems with the way computers affect the design process. The most notable problem is that the computer and the capabilities developed by programmers constrain the ways engineers can work. Rather than computers fitting into the ways engineers design, design is being adapted to the most obvious ways the computer can be used.

Chorafas (1987) discusses both the advantages of computer aids in improving engineering productivity and the limits of those aids. Engineers (Knight, 1983; Freeman, 1984) continually comment that the limits on design tools currently available make those tools at best an improvement to certain aspects of design. Genuine engineering design problems are too hard for computers to solve without supervision, although there are a number of applications where the computer can be used to aid an engineer.

With the prospect of realizable expansion of the computer into design, there came an emphasis on advancing design theory to the point it could be used as a basis for defining what portion of design can be allocated to the computer. However, the structured methods which were the most obvious uses of the computer led to an increased emphasis on structured versions of design theory, readily amenable to computerization. An early example of work toward more structured design theory is given by Purcell, Maller, and Gomain (1974).

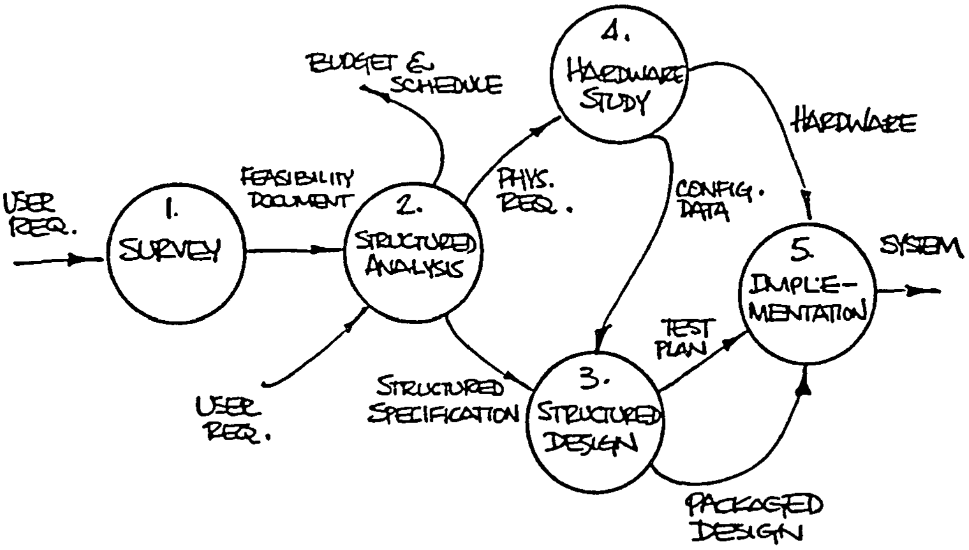
A later effort, reflecting the broader expanse of methods now available in a computer-based environment, is given in Tomiyama and Yoshikawa (1987), who attempt to describe the product development process in terms of set theory, with the solution lying at the intersection of all the requirements to be met. The structured descriptions of design serve as analogies for reasoning about the process, but they should not be misunderstood to be descriptions of a technique or procedure for accomplishing design.

Begg (1984) considers the particular problems of large, computer-based systems for design of integrated circuits. She expressed disappointment with the wide disparity between what is actually achievable in a practical working situation, and the apparent promise of recent research in the fields affecting design tools. Researchers have reported sweeping visions of improvements to be achieved in CAE which have yet to be realized. Michie (1977) acknowledges that some problems are too easy to warrant use of computers, while others are too hard for computer-based solution, which defines a somewhat restricted middle ground where problems are both easy enough to solve by computer and hard enough to warrant the programming effort. Researchers have dealt with rather simplified subproblems in many cases; mention was made in Chapter 1 of the "toy problem" in artificial intelligence research, such as a chess endgame, which has enough complexity to be interesting but is simple enough to be comprehensible.

An approach to making design problems clear enough to automate is to rely on hierarchical decomposition of the components. In this approach, each component is allowed to have only one function and components are isolated into a tree-structured hierarchy of functional subsystems. This type of structuring is the basis of Hubka's (1980) book on mechanical and structural design, Alexander's (1968) book on architecture and urban planning, and of De Marco's (1979) book on software design. But even in integrated circuit (Begg, 1984) and electronic system design (Finegold, 1984), emphasis on functional performance alone and isolation of individual components interfere with success. Medland and Jones (1984) report similar difficulties with application of functional reasoning in mechanical and electronic designs.

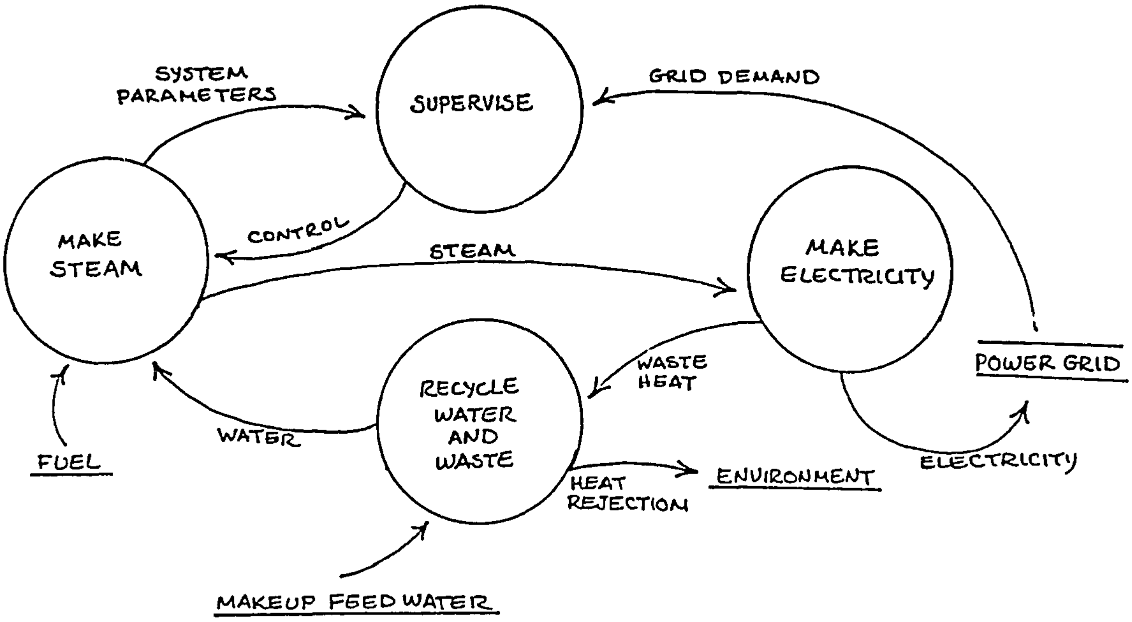
There is some value, however, in the philosophy behind hierarchical decomposition, often called "top-down" design. Such a design approach deals first with the overall functioning of the system, defining successive levels of greater detail in the context of prior decisions. The upper diagram in Figure 4 is the top level of a hierarchical description of the top-down software development process. The nodes in the diagram represent co-existent actors in the system, while the arcs represent the functional connections between those actors. Top-down design is a useful technique in software design (De Marco, 1979), and can be a valuable approach in other disciplines as well. The diagram in the lower half of Figure 4 indicates that in many mechanical designs, however, a purely top-down approach can mix many kinds of function together - energy streams and information streams, for example - while other issues of importance, such as reliability and structural integrity are neglected by the process because they are difficult to relate to the routine functional behavior of the system, which is what the top-down process deals with. The top-down design approach emphasizes function - performance - and neglects quality issues because they are not expressed in the same terms.

The author regularly uses top-down control over the design process, but it is important that engineers be willing to violate a strict top-down sequence when necessary. Often a few details will arise at critical points,



a software example of top-down design: software development process

(from De Marco, 1979)



a mechanical example of top-down design: electric power station

Figure 4

Examples of Top-Down Design

and then bottom-up design is necessary as those details drive several other decisions.

One of the earliest attempts to apply decision theory to design is the work of Starr (1963), who blended structured with intuitive methods of solution. Love's (1980) description of design was notable for two reasons. First, design theory was held to a pragmatic standard of whether it would serve the needs of practicing engineers. Second, the development of decision theory for engineering incorporated such "outside" issues as marketability and the cost of collecting evidence and making a decision. In other words, Love emphasized the engineer considering the whole product life when making a design decision.

Cross (1985) typifies the approach to decision theory for design, which deals primarily with the form the problems take and the strategies for solutions. There are, Cross argues, three key aspects of design thinking: intuition, resolution of problem situations (distinct from problem-solving, because a problem is often resolved by avoiding it entirely), and reasoning by analogy through the use of models.

Cross defined approaches to problem-solving which explain why a purely analytic, scientific approach to engineering can be a hindrance to design. Scientists tend to allow the form of the problem to lead them to solutions, which he calls a problem-focused strategy. This strategy is characterized by patience and a thorough examination of the problem, which succeeds only because time is a minor constraint. In contrast, engineers and architects proceed by proposing solutions and gradually eliminating some until a single solution remains, a solution-focused or product-oriented strategy. Engineers' problem-solving strategies are driven by limited time and budgets.

A growing branch of design theory emphasizes a philosophical basis for engineering, with a focus on the attempt to explore how decisions are made, instead of simply what knowledge is invoked to solve problems. Gero (1974) provides an early example of the attempt to deal with design primarily on a philosophical basis. Recognizing that many aspects of the whole problem of design were being neglected by the focus on those portions of design which could be structured, Gero dealt with the complete problem by invoking Gestalt or holistic principles.

Product development is dominated by holistic problems, that is, by concerns which span the whole project and problems which require the interaction of several different disciplines. Gestalt psychology is concerned with the behavior of the whole person (Wertheimer, 1959), and a Gestalt approach to design is intended to deal with the whole problem. When attempting to implement whole-problem solution approaches, the engineer can easily become lost in a flood of details. The top-down approach to design development is an attempt to deal with the whole problem by handling important portions at an abstract level, then deal with details in relative isolation after the interactions between primary issues have been defined.

Nadler (1981, 1986) typifies the broader approach to design theory, treating engineering problems as complete problems having both technical and social implications. In Nadler's more recent work, the study of design involves both technical issues (such as success criteria, knowledge representation questions, and the development of tools to aid decisions) and social issues (especially the psychology of decision-making, the economics which govern engineering solutions, the effects of tools on their users, and the underlying philosophy governing the industrial setting in which most engineers work). In Nadler's work, design theory is a philosophy dealing with the whole of the problem, rather than a set of methods for dealing with portions of design in isolation. This overall philosophy does, however, provide the means to develop detailed guidance on particular issues.

Nadler's approach to design may be described as "liberal" in the same meaning as the liberal arts or a liberal education. Emphasis is placed on recognizing the complex, multi-faceted nature of the problem, and the influence of several different subject areas when dealing with both the whole problem and whole solutions. There is a growing recognition that a broad, albeit incomplete grasp on the philosophy of design may be of greater practical value than clearly understanding any one aspect of the process.

To a limited extent, the philosophical basis for design has been present throughout the history of design theory. Asimow (1962) and Alexander (1968) combined specific structuring ideas with a broad philosophy about design. Their readers, however, took the structure and gave less attention to the philosophy. Thus Asimow is primarily remembered for the terms analysis and synthesis, which are frequently used to make a distinction between science content and design content in engineering curricula.

Winograd and Flores (1986), troubled by the strong trend to excessive structuring of solution methods and a narrowing, rigorous definition of rationality, cautioned that the design tool developer's prejudices and preconceptions have a strong influence on the behavior of the tool developed. Their approach to the design of computer-based tools applies concepts which had been of interest primarily to philosophers.

Koen (1985) provides an important contribution to the philosophy of engineering. In his monograph, Koen refutes the concept of rational design with the observation that engineers work in a changing, unpredictable, non-rational world, while rational approaches deal exclusively with unchanging situations. He puts much of the attempts to structure design theory into perspective, demonstrating both the uses and the limits of particular structuring approaches. He describes engineering as dominated by the use of heuristics, kernels of solutions whose validity depends on the context in which they are applied. All of design theory, he asserts, is heuristic. Based on both practical design experience and consideration of design theory, the author agrees entirely with Koen's description of the design process; however, he has an ample supply of detractors as well.

Because engineering design will always be slightly less than perfect in its achievements, there will always be room for improvement. It is important, however, that a change which looks good in the local context of a particular area requiring improvement does not have an adverse effect in the long run on the practice of engineering design as a whole.

Design theory today seems to be at an intersection. There is a path leading to design automation, in which increasing portions of design are carried out by computers under little or no human control. Methods based on both artificial intelligence and numerical modelling techniques are being explored. Paralleling the design automation path is one following the concept of structured methods of design which allow the design process to be more predictable. THe effort to make design more orderly is driven primarily by the wish to have a computer accomplish design. Predictability and control are appealing to managers who do not understand engineering design and are frustrated by the fact that prototypes don't always work correctly on the day the client visits.

Another available path deals with human decision theory. A better understanding of how humans make decisions is slowly becoming recognized as a key aspect of understanding how humans do design. At present, however, the decision theory path is not heavily travelled.

In another direction, a path follows the concept of rational design. Rational design is appealing because it bears the implication that all design work can be optimized. No extra material need be added to a structure which requires no factor of safety. This path is taken only occasionally, typically by an individual with minimal design experience.

There are other paths as well. The difficulty is that these paths are being viewed by most engineers from the viewpoint of an ant at the intersection. One can make out which directions are available, and how large a crowd is following each branch, but it is not possible to determine where each path leads. A bird flying overhead has a very different perspective of the same intersection. At this point in time, design theory needs a broad overview of the options. It is necessary to think about not only problems of contemporary interest, such as manufacturability and product quality, but also to anticipate future problems and consider how the solutions being developed today will affect future solutions. It is necessary to learn in general about improving design through the specific actions to improve design currently underway.

The author's use of the concept of quality of the design represents a unifying principle which can serve the need of showing how a choice of path might affect the future of design. A new idea for improving design methodology can be considered on its merits: does it lead to higher-quality designs?

3.2. Reasons to Pursue Design Research

A valid and important question is, why do design research? Often included, perhaps implicitly, is the observation that all the good design problems have already been solved. The author has shown some economic reasons for improving the design situation and has shown some of the avenues available to pursue in that research. This section will give the author's opinion on why a researcher might elect to pursue design problems, rather than some other subject of research.

An article in *Engineering Education* (Samuel Neaman Institute, 1987) indicates design is generally perceived by academics as "intellectually soft, intuitive, informal, and cookbooky". On the other hand, an "academically respectable" subject is "intellectually tough, analytic, formalizable". In these two phrases one observes the contradiction which lies at the heart of the misunderstanding of design.

Analytic, formalizable problems are closed, solvable problems. The problems have sufficient structure that it is clear from the outset that a solution exists, despite the fact that considerable work may be necessary to obtain a solution. A problem may require considerable mathematical or logical skill to solve, but the very fact it is formalizable limits its potential to stretch the mind. The knowns and unknowns all exist within some definable subset, allowing the problem to be approached in relative isolation. There are, however, a large number of research problems which are not directly formalizable, and researchers then apply substantial effort in developing an appropriate structure to impose during the search for a solution.

Now examine a typical design problem, which is not very structured. Design problems (and their solutions) are, in general, informal. In most cases there is not a clear separation between knowns and unknowns, or even between issues and non-issues. The information available to deal with the problem is the entire contents of the library. In many cases the solutions presented in the literature are for different problems, perhaps with some vague similarity. Especially in new technologies, the parameters as well as the knowledge representations required for design may be obscure or simply nonexistent.

The problem presented to the engineer is rarely complete. The first task is to develop a better definition of the problem. Once the solution steps are underway, it may be necessary to make assumptions, to rely on default values where needed values are unavailable, and to make other simplifications. At the same time, a large number of problems must be kept in balance. A partial list of the issues facing a designer include performance of the product, quality of the product, manufacturability, and product safety. The engineer is at the same time required to make effective use of design resources, which includes the engineer's own time.

To deal with these sticky, difficult, ill-posed problems in a swamp of only-partially-sorted information, the designer relies on such informal methods as judgement based on experience, and intuition. Thus it is appropriate to describe design as intuitive and informal. But design is by no means simply a "plug and chug" exercise using rules drawn from a handbook. Yes, many simplifications exist, but the simplicity of the form conceals the depth of understanding required to arrive at the simple method as well as the engineering judgement required to choose a valid simplification for a given problem.

Given a choice between a formalized, well-described, analytic problem and an informal, vague problem, researchers seem to prefer the former. A great deal of effort may be required to solve the analytic problem, but it is fairly certain from the outset that the solution is obtainable. More to the point, there is a single solution which can be verified as correct once it has been obtained. Such a problem may be difficult, but it is also relatively free from risk. In contrast, there is usually no single, correct answer to a design problem. Design work is full of risks.

Which problem, then, is the intellectually tougher one? It seems obvious that an ill-posed design problem, with no clear answer and no clear way of checking the answer's validity is more challenging. A researcher who wishes to grapple with very difficult problems would choose those which are only vaguely statable; which incorporate uncertainty, multiple solutions, and no clear way to check the validity of your solution; which require consideration of fields which may not be approachable from an analytic basis; and which require consideration of the interactions between multiple fields of study.

The author's interest in design does not derive entirely from a desire to seek out tough problems. Design has another attribute, which may contribute to its poor reputation among academics: it is a great deal of fun. A design problem is a puzzle, and like a very good puzzle, the solution is obtained through a combination of effort and good fortune. Design research is a meta-puzzle, in which it is necessary to examine the puzzle of how puzzles can be solved.

Chapter 4
Material Selection for Manufacturability

This chapter is a report on the author's exploration of a design problem which can be classified as decision making in the absence of clear information: accounting for manufacturability during material selection. Controlling the cost of manufacturing is considered a key to American industry's ability to continue competing in the world market. Manufacturability of designed parts is central to controlling these costs, and as a result design for manufacturability (DFM) has been given considerable attention. However, the approach to better accounting for manufacturing has taken the form of spreading the details of manufacturing out to be accounted for during design decisions.

The concept of simultaneous engineering has begun to take the place of DFM. To improve communication, simultaneous engineering places side-by-side the desks of the product designer and the manufacturing engineer who plans tooling and fabrication steps. One reason for simultaneous engineering is that the notion of DFM places responsibility on the designer to better anticipate the manufacturing engineer's decisions, which can add an unnecessarily heavy burden to the designer's list of concerns. Simultaneous engineering brings together product experts and manufacturing experts to deal jointly with two interconnected aspects of the whole problem, which is to get a profitable product on the market. However, simultaneous engineering requires the design engineer and the manufacturing engineer to have sufficient knowledge in common to work out their differences in the best interests of the design.

The question of putting a burden on the design engineer will be explored in this chapter. It is not necessarily inappropriate to expect the product design engineer to account for manufacturing in the design. What is incorrect is expecting the design engineer to deal with manufacturability in the same way as the manufacturing engineer would. It is reasonable to ask the designer to take responsibility for manufacturing, but these concerns cannot be handled at the same detailed level as during the planning and tooling process. Because the manufacturing planner deals with the particular details of fabrication, such as sequencing and allocation of shop resources and energy consumed during the steps, there is a substantial volume of both knowledge and data needed. It is unreasonable to expect the designer to deal with DFM at this level of detail. But the designer does have a general responsibility for a quality design, and broad economizing of the cost of fabrication is part of the quality which must be placed in the design itself. A broadly useful but economical version of manufacturing knowledge is needed to support design for manufacturability.

Where the specialist, such as a manufacturing engineer, deals in depth with the details of a particular issue, the designer is obliged to deal in breadth with the connections between many issues. For this reason it is not possible to expect the designer to deal with manufacturing, or any other single issue, at the same level as the specialist in that single issue. The general problem is to develop the necessary knowledge and information about each specialist area in a compact form which can be used in conjunction with knowledge about other issues to achieve a quality design. In this chapter, development of such knowledge is examined by the particular example of design for manufacturability during material selection.

A simplified, abstracted version of the knowledge used by manufacturing engineers is needed for use by designers. One such approach is a product announcement from a mechanical design tool company (Cognition, 1987): an expert system for cost of fabrication, the core issue in manufacturability. This approach still deals with the details, and is only capable of analysis of an existing design. In other words, the system is capable of providing feedback on manufacturability.

In this chapter the author will present manufacturability information in a form that allows it to be used directly, as a predictive, feed-forward aid to early decisions. The knowledge to be developed will take the form of general rules of broad validity. The methods discussed here support anticipating manufacturability during design decisions, rather than evaluation after the fact. This enhances economy by correcting errors before they can be made. In general one can reduce manufacturability concerns for metals to four areas: machining, forming, welding, and casting (Lindberg, 1964). Each of these will be considered in the context of material selection, a typical design decision.

A reader familiar with manufacturability issues will recognize that the development of manufacturing information presented in this chapter is merely an extension of much earlier work by others; work which is largely out of favor in modern manufacturing engineering because it lacks refinement. The purpose of this chapter is to show some structures appropriate to the design for manufacturability problem. Recognition and development of such structures will improve engineers' ability to carry out design as well as advance the state of design research. The bringing together of heretofore uncombined ideas also represents a worthwhile research endeavor. Material selection as a whole is largely misunderstood, and manufacturability in particular is completely ignored during the material selection process; a finding corroborated by the author's current co-workers. For instance, an engineer in the author's department has expertise in manufacturability and design for assembly as well as in engineering polymers through her experience in a company which makes molded plastic parts. She has found that most engineers fail to consider machinability when selecting the metal for an injection mold, despite the fact that mold-making is the single largest cost in a typical run of a part.

The author will demonstrate, through examples in dealing with manufacturability, some of the methods for reconstructing knowledge into forms which meet the designer's needs. The methods used will be mentioned where appropriate, and in the sixth chapter those methods will be discussed, drawing some general conclusions about the activity of knowledge restructuring.

4.1. Machining

There is a substantial volume of machinability data available for the manufacturing planner. One company, Metcut Associates, specializes in producing and indexing machinability data for particular combinations of metals, machining processes, geometries, and resulting surface roughness. Large industrial corporations have departments with similar responsibilities. In many cases this information is now available in the form of extensive databases.

Providing designers with the detailed machining information used by manufacturing planners is the idea behind integrated databases for design. An extensive manufacturing database will not necessarily aid an engineer in better anticipating manufacturability. There is simply too much detail for the designer to use it as an additional step in material selection. The depth of detail obscures the information which designers need most when selecting a material, which is an idea of whether candidate materials have comparable or widely differing machining costs. Because the information is not available in a form appropriate for searching out superior alternatives, it is not possible to search directly through machinability data for an appropriate material. Juvinall (1983) expresses concern the data can be unreliable as well. Data can be refined by further tests and greater detail, but refinement leads to still more detail, and does not help the design situation.

Designers are already balancing a large number of goals during material selection. To evaluate still another concern, the additional knowledge required must be available in terms reflecting the amount of effort available for that portion of the problem. An abbreviated form of knowledge is needed, combining an economically small evaluation with results which match reasonably well to the full description.

Such a form of knowledge about machinability is available. Datsko (1966) showed machinability can be described as a function of thermal conductivity, hardness, and ductility. The relationship is given:

$$v_{60} = C \frac{k}{H} \sqrt{1 - q} \qquad \text{(i)}$$

where v_{60} represents the cutting speed;

k represents the thermal conductivity;

H represents the Brinell hardness number;

q represents the reduction in area, expressed as a decimal fraction rather than a percentage; and

the coefficient, C, relates to the geometry of the cutting tool used, depth of cut, force on the tool, and other constants affecting the outcome, and a units conversion.

This gives the cutting speed for a tool life of sixty minutes, which can be used as a measure of machinability. Datsko compares experimental values with predictions from this formula and finds excellent correlations for routine engineering materials. The only poor prediction is for molybdenum and its alloys, which is a fairly uncommon material in mechanical engineering.

One doesn't estimate machinability to determine whether a material can be cut at all. Rather, the purpose of the evaluation is to compare the costs involved in obtaining identical parts from two different materials. Tool wear, the basis for the Datsko's version of machinability, is not the only consideration. Surface finish and surface integrity obtainable, how well tolerances can be met, and variability in the part produced as tool condition varies are all issues which affect overall consideration of machinability. Cook (1975) observes that these effects are all interrelated, and can be summarized through their relationship to tool wear. Thus a tool wear index is a reasonably valid indicator of all other issues in machinability.

Other variables might be used in a general expression for machinability. Ultimate strength could appear in the denominator, reflecting the fundamental fact that the tool cuts by exceeding a failure criterion to break away the removed metal. The strain-hardening exponent could be used, indicating ductility during plastic deformation, taking place between yielding and chip breaking, is an important influence on how the metal is cut. These two effects are the root causes behind most metal-cutting phenomena. In Datsko's equation, they are reflected indirectly, with hardness representing strength and reduction in area representing ductility. Datsko used material properties which were readily available, rather than properties which were perhaps more valid but harder to find for most metals. The use of readily-available material properties is an important choice, because the result is the calculation method developed is much more widely useful.

While Datsko's equation is not complicated, by noting a few basic features it can be further simplified for material selection considerations. Datsko's relationship is intended to give an actual estimate of tool life. A simple rating method has no requirement for direct physical meaning or units, so the coefficient may be dropped. For the same reason, the square root on the ductility term may be neglected without losing the monotonic relationship. Having abandoned any sense of units, hardness may be rated on any ascending scale, which includes any of the scales commonly used by engineers. Since elongation and reduction in area are both measures of ductility, elongation data may be substituted as well. The strain-hardening exponent is also an indicator of ductility, and may be used. Although assumptions such as these are not universally valid, for the ranges of values into which ductility and hardness data fall, they are reasonable.

This simplification yields a machinability index, similar in form to equation (i) but simplified in mathematical form and considerably more flexible in its application.

$$I_{m,1} = \frac{k}{H}(1 - D) \qquad \text{(ii)}$$

where k represents the thermal conductivity in any units;

H represents any hardness number such as:

- Brinell
- Rockwell (any scale)
- Vickers

D represents the ductility, measured by:

- reduction in area
- elongation
- strain-hardening exponent

The value (1 - D), then, can be considered a measure of the lack of ductility. The qualitative interpretation of equation (ii) is that lower-hardness materials are generally easier to machine, as are more brittle materials; this being supported by the conventional wisdom. It is necessary to be consistent in the use of the hardness and ductility measures, but only within a single comparison exercise. The measures could be different for another comparison, driven by the data which is available

A higher $I_{m,1}$ machinability index indicates a material which is easier to cut. This simple calculation offers a way to account for machinability with economy and reasonably valid ratings. It offers flexibility, which is appropriate since available data can take a wide variety of forms.

Further reductions are possible. For instance, if all the candidates have comparable thermal properties, that term may be dropped from equation (ii), leaving:

$$I_{m,2} = \frac{(1 - D)}{H} \qquad \text{(iii)}$$

The relationship between the two most important material parameters affecting machinability, strength and ductility, is made more visible by this reduction. More ductile metals are more difficult to machine because of the energy they absorb in plastic deformation prior to the chip breaking away. Higher-strength materials are more difficult to machine because of the greater energy required to exceed the breaking strength and remove metal.

It is well-established that mild steel is more machinable than the alloy steels. Machinability indices demonstrate the difference. The data involved in these calculations are shown in Table 1. Stainless steels are well-known for poor machinability, and this is reflected in the index computed in Table 1. The table also compares the $I_{m,1}$ machinability index with Datsko's V_{60} measure, demonstrating the desired indication of relative machinability is preserved despite the simplifications made in the derivation of the index.

Table 1
Material Properties and Machinability for Three Steels

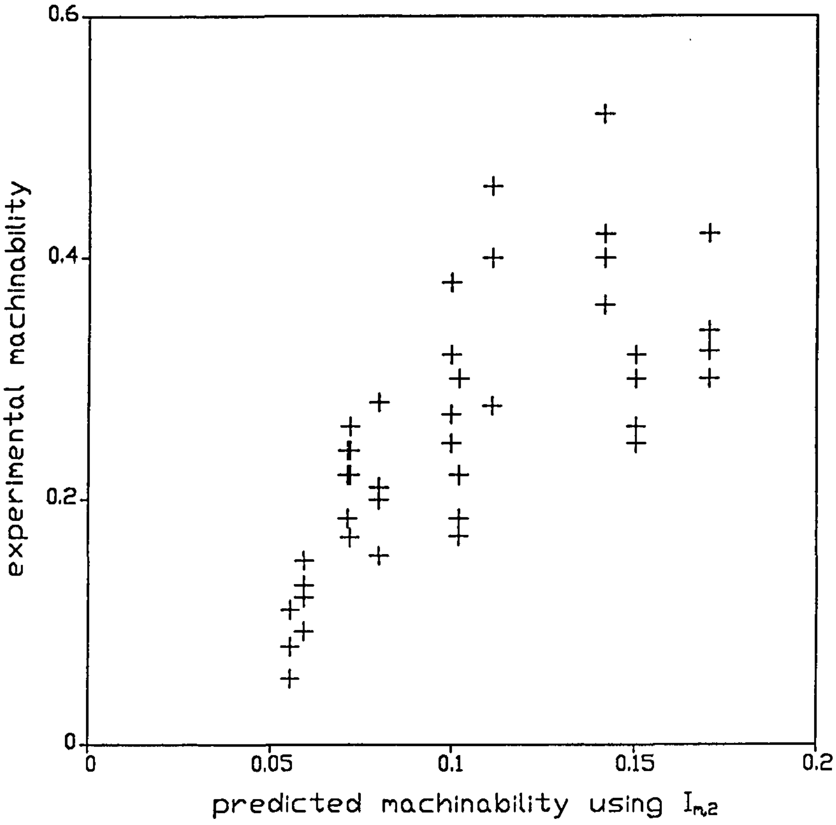
material	hardness	ductility	k	$I_{m,1}$	V_{60}
1020 steel	143	59%	52	0.149	268
4340 steel	206	43%	38	0.105	167
304 stainless	266	62%	16	0.023	42.7
units	Brinell	RA%	W/m°K		

Another example of the use of this index for design is deciding how to sequence machining and heat treatments. To repair a Charpy impact test machine, the author made new latch plates for the locking mechanism. 1045 steel was chosen as the material, and in the machine the component was to be normalized. The bar stock from which the parts were to be made is hot-rolled, and it appeared time could be saved by heat treating the stock before cutting the final part, suggesting the question of whether heat treating before or after machining would be appropriate. Table 2 shows the material properties in question, and the index values obtained. It was decided that the hot-rolled condition was substantially more machinable, and therefore heat treating before machining was not warranted. In this case, the dimensional changes which always result from heat treatments were not large enough to make a difference. Should exact dimensions be important in a component, the engineer can also reason about how much machining after heat treatment is appropriate by comparing the machinability before and after heat treatment.

Table 2
Machinability for 1045 Steel Component

treatment	hardness	ductility	$I_{m,2}$
hot-rolled	163	40%	3.7
normalized	193	48%	2.7
units	Brinell	RA%	$x10^{-3}$

The index discussed here correlates reasonably well in Figure 5 with machinability test data under a variety of conditions. The magnitudes of the indices being compared differ because the experimental value has been reduced to relative machinability from a variety of sources, and the predicted value is the index discussed above. The test data in Figure 5 includes drilling, side-milling, and lathe-turning tests, and incorporates such measures of machinability as tool life, tool breakage, force on the cutting tool, and power consumption. The correlation is demonstrated by the two different measures of machinability remaining monotonic and fairly linear.



experimental data sources:

Abeyama, Kimura, and Nakamura, 1983; Araki, et. al., 1975; Zlatin and Christopher, 1975

Figure 5

Predicted and Experimental Machinability

The machinability index presented above is necessarily an estimate. It produces values for comparison, but the values must be interpreted. In particular index values for the candidate materials cannot be taken simply as a rank ordering system. If the values are spaced some distance apart on the scale, then such ranking is not unreasonable. But if the values fall near one another, order becomes meaningless and one must simply conclude the materials are nearly equivalent under the parameter. In material selection, either of these two interpretations provides useful information. It can be argued the situation is no better when using detailed machinability data.

A more exacting approach could compile machinability data for classes of material, compressing the voluminous data books into a few pages of information. But the simple formula has demonstrable value over even a condensation of the detailed information, because it relies on data already available in the standard material references, and because the index equation itself gives the designer insight into how the rankings are established and what properties characterize superior alternatives. The equations given above constitute direct, predictive knowledge on material selection for machinability.

In general, the engineer is not searching for the single clear choice when checking machinability during material selection, rather a reasonable material is sought which meets a number of criteria. In some cases all the candidates are difficult to machine; this is something the designer should become aware of as materials are considered. It is not always possible to choose a material which is easy to process through the manufacturing steps;

but when the material is difficult to process, it should be the result of a conscious decision during design instead of an unpleasant discovery during manufacturing.

The machinability indices discussed above, and the way they were both derived and applied, illustrate three important points about figure-of-merit calculations. First, although the reasoning behind the index dealt primarily with tool wear, other phenomena are equally represented (Cook, 1975). The data compared in Figure 5 include several other measures of machinability which also correlate well with the index. The simplicity of the final form conceals considerable depth of knowledge about the phenomena involved.

Second, the index is not a simple "cookbook" formula which can be applied blindly to make decisions by a thoughtless rule. Rather it represents an economical way of developing evidence for the basis of a decision. Usefulness depends on the user fully understanding the basis for the index, and in fact it is quite easy to mis-use a figure-of-merit comparison such as the index by failing to fully comprehend its use. However, the very act of using the index is a mnemonic aid, reminding the engineer to weigh the evidence carefully.

Third, simplifying the equations to their barest form serves two useful purposes: the relationship between the variables in the equation and the underlying physical effects is made clearer, and the lack of precision in the calculation is more prominent. Lack of precision in indices used for early design decisions is a virtue, deliberately created, as a reminder to the engineer of the inexact nature of the comparison being made.

4.2. Forming

Forming processes deal with changing the shape of the material through inelastic deformation. The mechanics of plastic deformation are reasonably well understood in general, but are also rather complicated to apply to a particular problem. A wide variety of mechanical components are forged, including engine crankshafts and connection rods, and aircraft struts and structural components. Pressure vessels and sheet-metal products rely heavily on cold sheet-bending.

During material selection, the engineer's concern is simply to assure a reasonably formable metal is chosen from the list of candidates. Thus it is not necessary to evaluate the particulars of the planned forming operation, it is only necessary to consider which of the candidates is more easily formed in a generic sense. This involves two questions: the energy (or force) required to form and limits on forming. Both of these concerns can be simply related to basic material properties. In the discussion to follow, formability will refer to the general limits on forming, while forgeability will refer to the general limits on forming at hot work temperatures.

For both open and closed die forgings, the required ram force (Dieter, 1986) is

$$P = C\ S_0\ A \qquad \text{(iv)}$$

where S_0 is the flow stress;

A is the area enclosed by the parting line of the dies; and

C is a correction factor which accounts primarily for the geometry of the part outside of the plane of the parting line

To achieve uniform flow in the die, it is necessary to bring the entire part above the flow stress. For a given forging, C and A are both constant. Thus to compare the forces required to form two different candidate materials it is only necessary to compare yield strengths. Wick (1984) also makes the observation that formability is generally proportional to the inverse of the yield strength.

In sheet and plate, bend radii are usually dealt with in terms of the plate thickness. The limit on bend radius is primarily associated with the maximum allowable strain on the tension side. The geometry of plate bending is diagrammed in Figure 6. A 3T bend refers to a bend radius, R, three times the plate thickness, h; for a 3T bend C equals 3. Therefore, for a given bend,

$$C = R / h \qquad \text{(v)}$$

Datsko and Yang (1960) defined C_{min} as the minimum ratio between bend radius, R, and plate thickness, h, for which tearing does not occur, defining the smallest feasible bend radius. The ratio C_{min} varies from metal to metal, and increases with prior cold work. C_{min} can be related to the reduction in area, q, for a simple tensile test:

$$C_{min} = \frac{1}{2q} - 1 \; ; \qquad q < 0.2 \qquad \text{(vi)}$$

and

$$C_{min} = \frac{(1 - q)^2}{2q - q^2} \; ; \qquad q > 0.2 \qquad \text{(vii)}$$

Almost all materials which are serious candidates for forming have values of q greater than 0.2 (20 per cent reduction in area) and equation (vii) will apply. For these ductile materials, there is greater reduction in the tension side of the plate in bending than on the compression side. Greater ductility shifts the neutral axis, leading to the more complicated expression. The second equation is conservative for lower values of reduction in area.

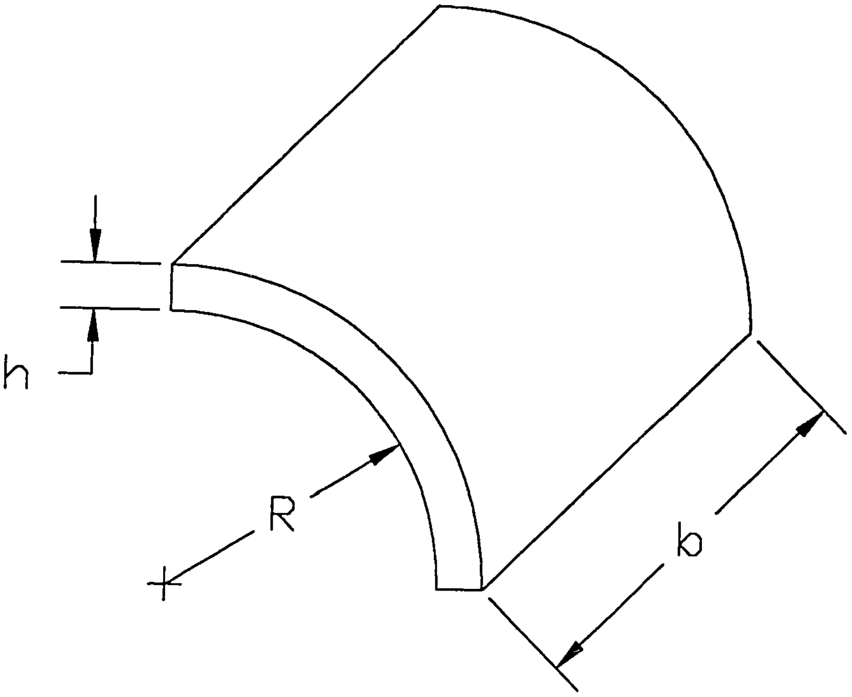


Figure 6

Geometry of a Plate Being Bent

These two equations illustrate an important property of heuristics. There may be multiple, contradictory solutions to a given problem, as in these two equations. But the value of a heuristic is not absolute, there are limits on its validity. In the example above, the limits are provided both explicitly, in the 20% division between equations, and implicitly, in the general bounds on values of reduction in area for metals.

There is a biaxiality effect, dealing with the shift in constraint from plane stress to plane strain along the axis of the bend. For a width of plate, b, along the axis of the bend, when b/h is small there is essentially a plane stress situation. As b/h increases, plane strain becomes the condition away from the edges, and a second principal stress is induced along the axis of bending. This effect has some limits, and for a value of b/h greater than 10 the strain to cause tearing may be reduced to a minimum value of half the plane stress value. But for most cases of interest, b is less than one inch and h is greater than one foot, so b/h is always greater than ten. The equations above reflect this effect. For the purposes of making a comparison between two metals they are valid regardless of the ratio b/h, because for a given geometry the effect is accounted for by a constant factor (Datsko and Yang, 1960).

The yield stress of a metal can be used as an indicator of the force required to form a component. But as the C_{min} assessment above indicates, most forming difficulties are associated with excessive strains, not with excessive stresses. The maximum allowable strain, in general, is the strain at the onset of unstable plastic deformation, which is in a tension test the point where necking begins. An estimate of the limiting strain can be used as a figure-of-merit for comparing formabilities of candidate materials.

An estimate of the maximum allowable strain during plastic deformation is needed. This is indicated by the strain at the onset of necking. Taking the theme developed in the preceding section, the problem is to relate this strain to the material properties which are readily available. These properties are yield stress, ultimate strength, hardness, reduction in area, and elongation. There is a relationship between yield and ultimate strengths which can be used to determine strain-hardening, since strain hardening accounts for the rise in flow stress from initial yielding to the onset of plastic instability at the ultimate strength.

Theory predicts the true plastic strain at the onset of plastic instability (the point at which necking takes place in a tensile specimen) is equal to the strain hardening exponent, n. This provides a point of departure, since there is a direct relationship between work hardening and the elevation of the flow stress from yielding to ultimate strength. Starting with the basic relationship between stress (σ) and strain (ϵ):

$$\epsilon = \frac{\sigma}{E} + \left[\frac{\sigma}{K}\right]^{1/n} \qquad \text{(vii)}$$

and at necking, the plastic component of strain, ϵ_p, is approximately equal to the strain hardening exponent

$$\epsilon_{neck} \approx \epsilon_p = n \qquad \text{(viii)}$$

The second term of equation (vii) gives the expression for the plastic component of strain. By substitution one can obtain

$$\epsilon_{neck} = \left[\frac{\sigma}{K}\right]^{1/n} \qquad \text{(ix)}$$

The true stress at necking is equivalent to the engineering ultimate strength. Engineering (S) and true stresses (σ) are related by the engineering strain:

$$\sigma = S\ (1 + e) \tag{x}$$

in general, and at necking the same relationship holds between the true stress at necking and the ultimate tensile strength

$$\sigma_{neck} = S_{UTS}\ (1 + e_{neck}) \tag{xi}$$

So from (ix) and (xi) one obtains

$$n = \left[\frac{S_{UTS}\ (1 + e_{neck})}{K}\right]^{1/n} \tag{xii}$$

The relationship between engineering (e) and true strain (ϵ) is given by

$$\epsilon = \ln\ (1 + e) \tag{xiii}$$

Then the true strain at necking, approximated by the strain hardening exponent, n, can be equated to engineering strain at necking

$$\epsilon_{neck} = \ln\ (1 + e_{neck}) \approx \epsilon_p = n \tag{xiv}$$

$$\exp(n) = 1 + e_{neck} \tag{xv}$$

$$e_{neck} = \exp(n) - 1 \tag{xvi}$$

Equation (xvi) can be inserted into (xii), to obtain

$$n = \left[\frac{S_{UTS}\ (1 + \exp(n) - 1)}{K}\right]^{1/n} = \left[S_{UTS}\ \frac{\exp(n)}{K}\right]^{1/n} \tag{xvii}$$

which reduces to

$$n^n = \frac{S_{UTS}}{K} \exp(n) \tag{xviii}$$

The coefficient of the plastic portion of stress, K, remains to be eliminated. Using the 0.2 per cent offset definition, the plastic strain at yielding is 0.002. Then

$$0.002 = \left[\frac{\sigma_{YP}}{K}\right]^{1/n} \qquad \text{(xix)}$$

True stress at yielding is nearly equal to engineering yield stress, giving

$$0.002 = \left[\frac{S_{YP}}{K}\right]^{1/n} \qquad \text{(xx)}$$

$$\frac{1}{K} = \frac{0.002^n}{S_{YP}} \qquad \text{(xxi)}$$

Substituting (xxi) into (xviii)

$$n^n = \frac{S_{uts}}{S_{yp}} \, 0.002^n \, \exp(n) \qquad \text{(xxii)}$$

For convenience, define the work hardening ratio, r

$$r = \frac{S_{uts}}{S_{yp}} \qquad \text{(xxiii)}$$

then

$$n^n = r \, 0.002^n \, \exp(n) \qquad \text{(xxiv)}$$

$$r = \left[\frac{n}{0.002 \, \exp(1)}\right]^{1/n} \qquad \text{(xxv)}$$

$$r = (184 \, n)^n \qquad \text{(xxvi)}$$

Given a strain-hardening exponent, n, it is possible to predict the ratio between yield and ultimate strengths. But the goal was to solve for n, given a value of r computed from tabulated values of ultimate and yield. Equation (xxvi) cannot be solved in closed form for n, but plotting the curve of the equation in Figure 7, one can see that, given a value of r, there is a single value of n which corresponds. Solution by iteration, for example, should converge on the correct solution.

An extension of this approach is to develop an approximate solution. As a first pass, a simple linear fit was attempted, using least squares. For the range of greatest interest,

$$0.05 < n < 0.4 \qquad \text{(xxvii)}$$

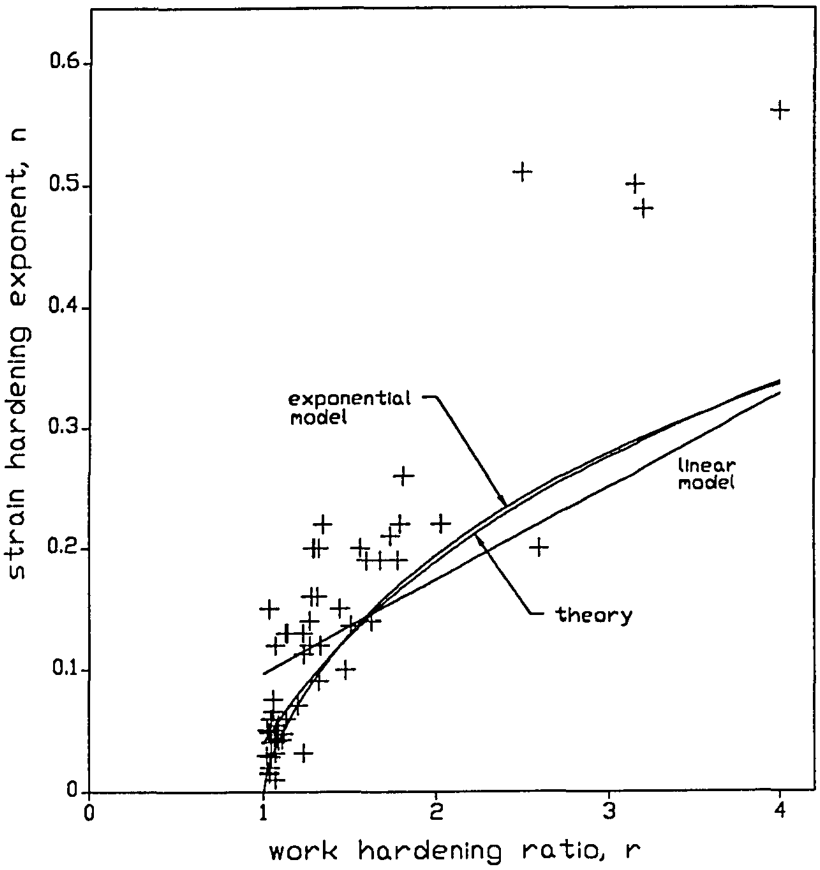
even the linear correlation comes within about ten per cent of theory. The linear relationship obtained was:

$$n \approx 0.020 + 0.077\ r \qquad \text{(xxvii)}$$

Reasoning that an equation of the form x^x should have a form similar to the exponential function e^x, a fit to this form was attempted. The exponential approximation, also developed using least squares fit on the values predicted by theory, fits theory with an error less than about four per cent for the range of exponents of practical interest for forming operations. The exponential approximation is:

$$n \approx 0.040 + 0.215 \ln (r) \qquad \text{(xxix)}$$

Both the linear and the exponential approximations are shown in Figure 7. The theoretical and approximate curves are equally poor in comparison to the data showing experimentally-determined values, but the heuristics above are the best available method which offers any economy. The



data source: Tucker, Landgraf, and Brose, 1974

Figure 7

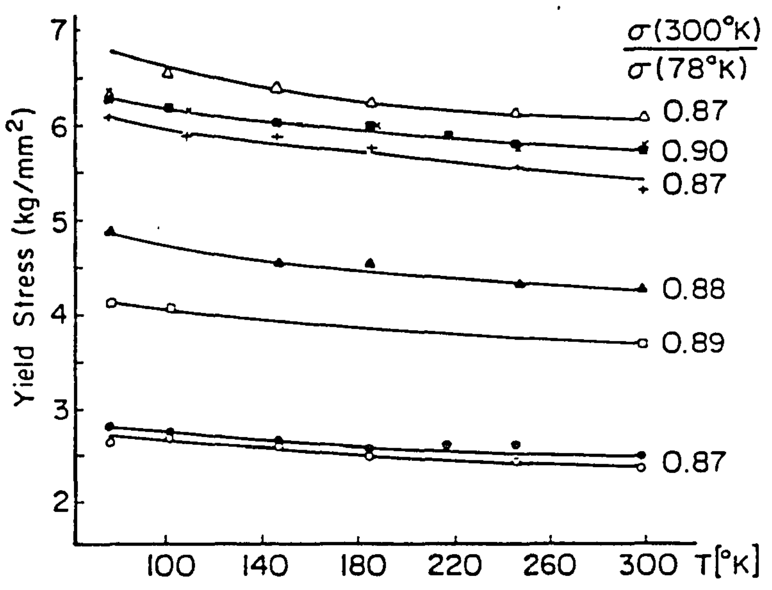
Comparing Theoretical and Two Prediction Models

of Work Hardening Exponent to Experimental Data alternative is to search for strain-hardening exponent data, which is difficult to find in the literature for any range of materials.

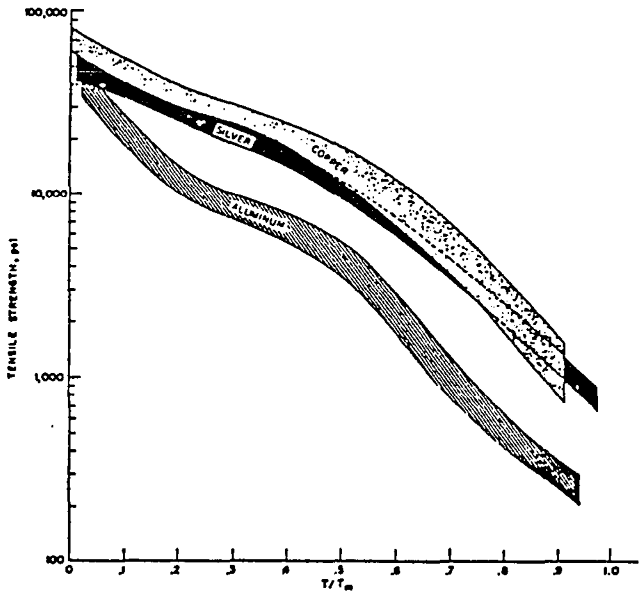
This derivation illustrates a heuristic frequently used by engineers: adapt theoretical, analytical information into forms which provide insight into the problem of interest.

The forming temperature is also an influence on the overall cost of the part. Representative metals have reasonably identical relationships between flow stress and temperature, (see Figure 8) thus rank ordering by flow stress at room temperature (the yield stress from a tensile test) would remain valid for warm work temperatures (cold work at elevated temperatures), or hot work. There are temperature limits on hot work, since the temperature must be in the creep range, that is, above one-half the melting point (absolute temperature scale), setting a lower limit on temperature. The upper limit, especially in aluminum alloys and high-carbon steels, is the eutectic temperature, since above that temperature a portion of the forging will be liquid metal. Presence of liquid phases during forging results in considerably less ductility, which is known as hot shortness. Oxidation at elevated temperature is also a problem for some metals, and may place an upper limit on forging temperature as well. Selection of proper hot-work temperature, however, is not a major factor in considering forgeability, which is formability under hot-work conditions.

Actual variation in the flow stress with temperature is more complex than the assumption made above. But from the evidence available, the assumption is a reasonable one. The fact that the data in Figure 8 deals with metals which are not the most common in mechanical engineering illustrates an important design heuristic: if the "right" data isn't available, adapt the



(from Grosskreutz and Mughrabi, 1975)



(from Hosford and Caddell, 1983)

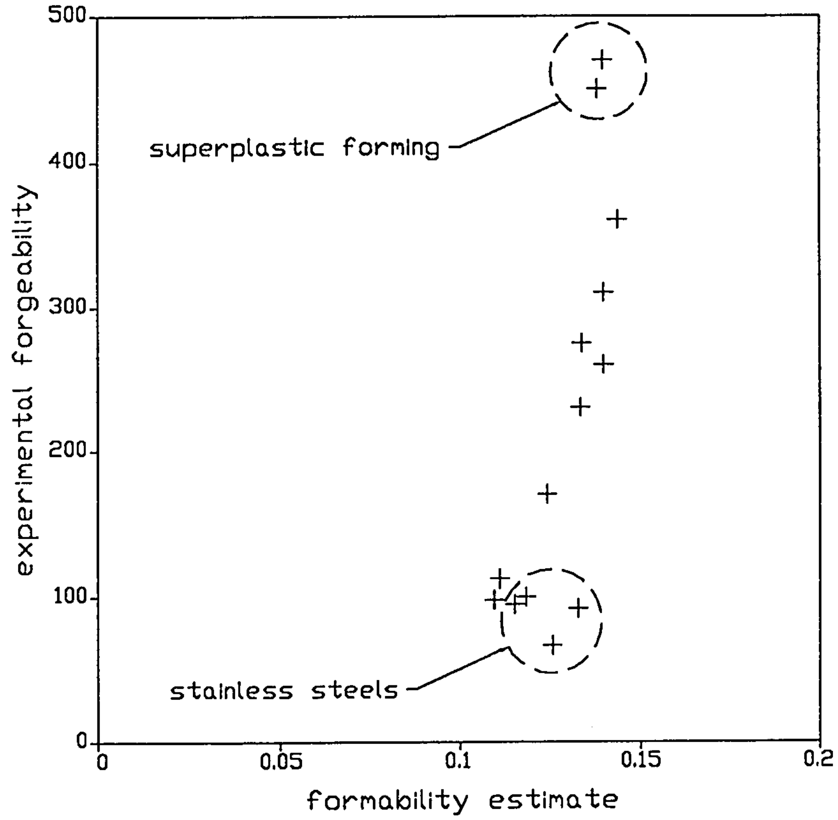
Figure 8

Variation in Flow Stress with Temperature "wrong" data to guide reasoning about the problem. The basis for the assumed relationship between the metals shown in Figure 8 and metals of practical engineering interest is the fact that temperature effects are driven more by the general thermodynamics of metals than they are by individual alloy properties, and the metals in Figure 8 incorporate an adequate range of crystalline types to cover metals of engineering interest.

The notable exception to the assumption of consistent high-temperature behavior is the austenite transformation in ferritic steels. In general it is necessary to raise a steel above the austenitization temperature to achieve hot work temperatures. It will be shown later in this section, however, that the high-temperature transformation to austenite does not interfere with useful estimates based on room-temperature data.

Strain-hardening exponent also varies with temperature. But just as the flow stress variation is driven more by general properties of materials than by the particulars of an alloy, the strain hardening exponent's variation seems to be consistent enough that room-temperature comparisons are reasonable for predicting elevated-temperature behavior.

A comparison of predicted forgeability, using room-temperature test data in the exponential approximation above, with experimental forgeability data at hot work temperatures from Timken (1974) is given in Figure 9. The forging data is from hot twist tests, and the vertical axis of Figure 9 gives twists to failure at the recommended forging temperature. On the horizontal axis, the predicted formability is the linear-fit equation for the work-hardening exponent. The magnitudes of these two different measures of formability differ, but the data shown in the figure indicates the author's estimating method, although not a strong method of differentiation, is good enough to be



data source: Timken, 1974

Figure 9

Predicted and Experimental Forgeability

useful for comparing relative formability of candidate materials. Reasonable prediction of forgeability was obtained for plain carbon and alloy steels, and for nickel superalloys. There are materials for which the room-temperature approximation overpredicts forgeability, the stainless steels, but all stainless steels for which data was available were comparable to one another and not as formable as other steels, due to the high content of alloying elements. Except for the stainless steels, the figure indicates reasonable correlation between hot-twist forgeability test results and the figure-of-merit approximation of the strain-hardening exponent developed in this research.

The geometry of the formed part has a strong influence on its formability. But less-formable materials generate limits on the geometry. As noted above in the discussion of sheet bending, the limiting radius is governed by ductility, a material property. For complex shapes, it is customary to carry out a series of formability tests, such as those required to obtain a Keeler-Goodwin forming limit diagram (Hosford and Caddell, 1983). The purpose of the discussion above is to provide enough information to the engineer that some preliminary decisions can be made in the absence of such data. From this basis, one can also decide whether the experimental program required to obtain the forming limit diagram is needed and which materials are best suited as candidates.

For the purposes of evaluating formability during material selection, it is necessary to consider two indices: the flow stress, a measure of the force required to form the part; and the strain-hardening exponent, a measure of the limits on deformation. These indices are based on three tensile test parameters which are readily available: the yield strength, the ultimate strength, and the reduction in area. For the purposes of preliminary decisions about material choices, one can neglect temperature effects and still obtain reasonable comparisons of formability.

This exercise is reasonably useful in itself, but its primary value is as a demonstration of the thinking required as an engineer develops a problem, connecting information which is needed to data that is readily available. In this particular situation, the formal structure provided by theory did not supply a convenient way of determining the parameter desired. The data from testing shown in Figure 7 shows theory to be an approximation of reality; other approximations represent the data equally well, and are easier to manipulate to support design. Two ingredients were used in creating another, more convenient approximation. First, it was necessary to restrict the range of the approximation to a range of particular interest. Second, it was useful to reason about the form the approximation should take, based on the form of the equation to be replaced. These two heuristics are examples of the methodology which can be used to obtain more economical representations of analytic knowledge.

4.3. Welding

Welding is problematic because the involved processes are complex, requiring application of virtually every major topic in engineering, including heat transfer, solidification structures, microstructural effects of various heating and cooling cycles, magnetic and electric effects on molten metals, electric power, and in some cases combustion or plasma physics. As a result there seem to be two classes of engineer: those who know nothing about welding, and those who are welding engineers. Both groups treat welding as a complex process which requires evaluation in great detail. This generalization overstates the situation, but welding knowledge has yet to achieve wide availability in a form which has usefulness in design.

For the discussion above on machining, the approach was to summarize into a single parameter, easily derived from available material data, the effects which must be considered. In welding, it is necessary to account for a number of possible problems. Alloy composition is the influence on many of these modes, providing the means to account for them as the material is chosen. A volume of detailed information is available, but a condensed form can adequately account for many aspects important during early design.

Three ways of reasoning about aluminum weldability approach the same information from different directions. Weldability of aluminum alloys can be anticipated very simply (Alcoa, 1960). Pure aluminum lacks the strength required in most applications. Two strengthening mechanisms are used in aluminum alloys: precipitation hardening in the 2000 and 7000 series alloys, and work-hardening in the 1000, 5000, and 6000 series alloys. Work hardened alloys, without exception, are weldable with moderate reduction in the strength through the weld. Precipitation hardening is achieved by careful control of both temperature and time through a heat treatment sequence. One can apply the heuristic mentioned in the next chapter, "every heat treatment is present in the weld", to reason that a precipitation-hardened aluminum alloy will be adversely affected by the heat treatments due to the weld. Precipitates are coarsened or redissolved in the vicinity of the weld, leading to loss of strength and in some cases loss of ductility as well. Some precipitation-hardened alloys will crack at the weld if a second solution heat-treat is attempted. In general, precipitation-hardened alloys are not weldable. There are a few exceptions, such as 2014, which are weldable but require a carefully controlled, low-heat-input process which involves greater expense that is normally assumed for welding.

Weldability division by alloy number follows the chemistry of the alloys. Copper, zinc, and magnesium-zinc alloys of aluminum are precipitation hardened and are not weldable. Magnesium and silicon-manganese alloys of aluminum are weldable. Silicon alloys, which are used primarily for casting, have decreasing weldability with increasing silicon content. This information about the effects of alloying elements on weldability will help an engineer to consider weldability of aluminum alloys which are not given standard wrought-aluminum alloy numbers, such as castings.

One additional method for inferring weldability relies on the temper designations following the alloy number. Alloys with a "-T" temper designation (2024-T3, 7075-T6) are precipitation hardened. Work hardened alloys have a "-H" designation (5052-H38). Thus alloys which have a "-T" temper designation are not, in general, weldable; alloys with "-H" temper designations are weldable. By examining the alloy number or the temper designation of an aluminum, the engineer can determine if the alloy is work-hardened or precipitation hardened, and from that information infer weldability. Weldability can also be inferred from the chemistry.

For steels, there are two basic welding problems to be considered during material selection: cold cracking and hot tearing. These problems can be dealt with by fairly simple means. Several other forms of welding problem can be related to the same root causes (Rogerson, 1983), including reheat cracking and chevron cracking. Lamellar tearing is a significant weld problem which is dealt with primarily in the geometry of the welded joint, and for that reason is the only significant source of welding problems which is not considered in this section.

Cold cracking is an embrittlement problem (Graville, 1975). Although the problem is related to hydrogen in the weld metal, one can simply imagine the weld is brittle because it has cooled too fast. This analogy has merit since the tendency to cold crack is strongly related to hardenability: greater hardenability means greater problems with cold cracking.

Hardenability in turn can be summarized by the carbon equivalent (Heuschkel, 1949), which is a fairly simple calculation. Each element which contributes to hardness can be treated as an equivalent amount of carbon, allowing an alloy steel to be reduced to a equivalent plain carbon steel. There are several forms of the carbon equivalent in the literature, often contradicting one another in some of the details; Blodgett (1985) uses the most common version:

$$C.E.= \%C + 1/6\ (\%Mn+\%Si) + 1/5\ (\%Cr+\%Mo+\%V) + 1/15\ (\%Ni+\%Cu) \qquad (xxx)$$

There is detailed information available on the interpretation of this value in the context of other variables such as plate thickness (for instance, Heuschkel, 1949). For a carbon equivalent greater than about 0.5%, cold cracking will be a problem and should be considered as the material is selected.

To consider the weldability of candidate structural steels - say, ASTM A-36, A-516, A-529, and A-588 - it is necessary to compute the carbon equivalents, which are shown in Table 3.

Just because a material is a candidate for cold cracking, it is not necessary to reject it as a candidate for the weldment. A number of steps can be taken to weld a steel with a high carbon equivalent. The designer can anticipate there will be increased welding costs in a material with a high equivalent. Cold cracking in a weld requires heating with rapid cooling, a hydrogen source, and constraint on the weld to add residual stresses. Hydrogen may be limited by cleaning and better inerting of the weld pool. Low-hydrogen practice involves higher quality welding and might also involve a more expensive welding process. Constraint may be reduced in many cases by the design of the weld joint. Although the heating is unavoidable, preheating the weldment will slow the cooling rate, and a postweld stress relief heat treatment will both reduce residual stresses and allow hydrogen to diffuse back out of the weld. These represent ways to deal with poor weldability, but at the point of material selection it is sufficient to know what weldability problems may arise. Low weldability can be anticipated and planned for, rather than an unpleasant surprise later, when production is underway. The carbon equivalent, and other methods discussed here, illustrate how an economical calculation when the material is originally chosen can warn about the possibility of a manufacturing problem which may otherwise not be discovered until correction has become very expensive.

Table 3
Carbon Equivalents for Four Structural Steels

ASTM alloy	carbon content	carbon equivalent
A-36	0.26%	0.43%
A-516	0.24%	0.45%
A-529	0.27%	0.48%
A-588	0.20%	0.60%

In plain carbon and alloy steels, hot tearing is related to the presence of liquid phases of low melting point, which wet grain boundaries. Note the strong similarity to hot shortness, discussed in the earlier section on forming. Because the usual source of liquid phases is low-melting-point compounds such as iron sulfide, control of hot tearing is related to good control of the sulfur content. As a rule, the manganese content should be 35 to 45 time the sulfur content for typical weldable carbon steels, and in fact this is usually provided for by standard steelmaking practice. This approach succeeds because sulfur prefers manganese, forming a higher-melting compound which, in the solid state, takes a spherical forms which limits the effect of the inclusion. In greater detail, the level of manganese required depends on carbon - the minimum manganese content increases with increasing carbon content. For the range of carbon contents which are routinely welded, however, the proportion reported above - corresponding to roughly 1 to 1.5 percent manganese - is a reasonable one. Manganese contributes to the carbon equivalent, however, so it is necessary to make a compromise in some cases.

In stainless steels, stabilized austenite is actually a problem when welding. Hot tearing reflects local failures as the weld cools, due to large thermal strains and low strengths at very high temperatures. Fully austenitic steel has a strong tendency for this kind of tearing because of austenite's poor strength and poor ductility at high temperature and the slow cooling rate due to low thermal conductivity. But a minimal level of ferrite in the weld, about five per cent, can suppress this problem, although the exact reason why this is so is not clear.

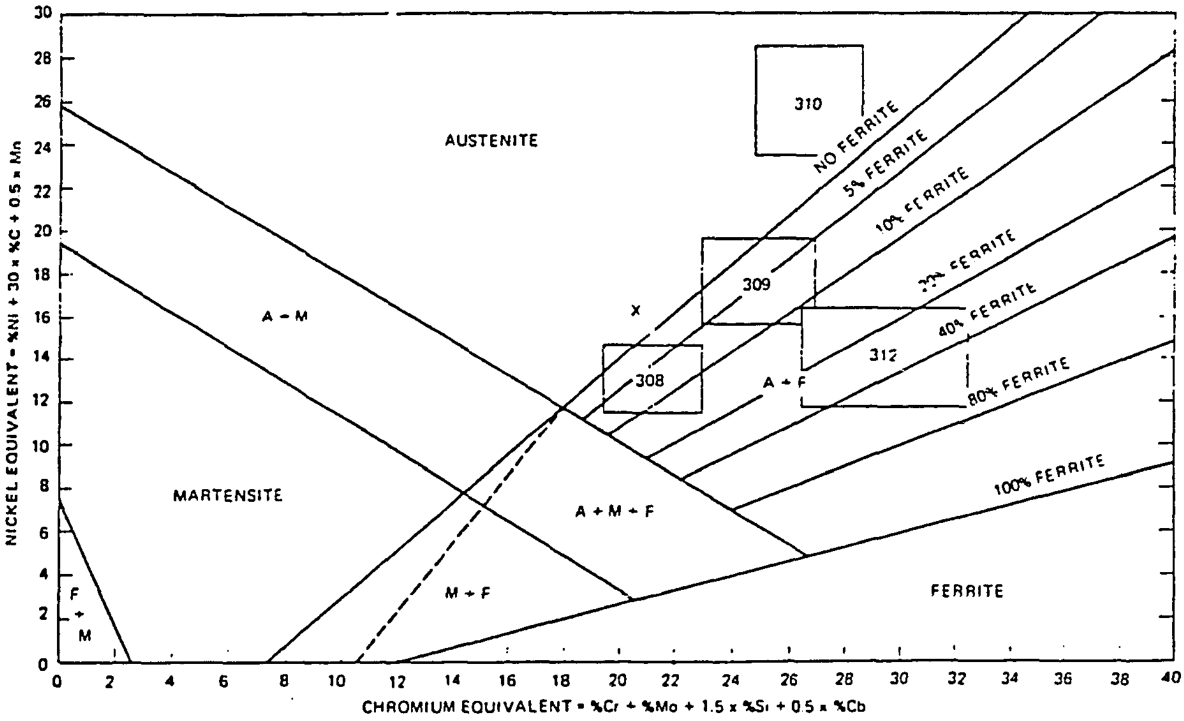
To predict the phases obtained it is necessary to evaluate the balance between elements which stabilize austenite and elements which stabilize ferrite. The first group is related primarily to nickel in the stainless steels, the second to chromium. Other elements can be included as equivalent amounts in the same manner as elements affecting hardenability can be related to the carbon equivalent:

$$\text{Ni.E.} = \%\text{Ni} + 30\ \%\text{C} + 1/2\ \%\text{Mn} \qquad \text{(xxxi)}$$

$$\text{Cr.E.} = \%\text{Cr} + \%\text{Mo} + 3/2\ \%\text{Si} + 1/2\ \%\text{Cb} \qquad \text{(xxxii)}$$

One then turns to the Schaeffler Diagram (Schaeffler, 1949; the diagram can also be found in the annual Databook Issue of *Metals Progress*), which predicts the phases present from these two equivalents. The best-known version of the diagram is shown in Figure 10. The diagram is not exact, but is a useful rough estimator for fast, economical early evaluations. The Schaeffler diagram is no longer considered exact enough for welding engineers, who need more detailed information as they look at the particulars of a problem. The diagram is, however, a useful estimator for fast, economical early evaluations. In essence, time-temperature transformation diagrams for a number of different compositions have been summarized for the cooling rates typically encountered after the weld is made. Welding processes developed in recent years, such as laser, plasma arc, and electron beam welding, involve lower, more localized heat inputs. This leads to faster cooling rates and generally leads to improved weldability, which affects the accuracy of the Schaeffler diagram. A fairly straightforward welding research effort could modify the diagram to account for newer welding practices, but welding researchers do not consider the Schaeffler diagram worthy of further development.

Anticipating weldability of steels during material selection requires consideration of indices which are based on alloy chemical composition. The



Example: Point X on the diagram indicates the equivalent composition of a type 318 (316 CB) weld deposit containing 0.07 C, 1.55 Mn, 0.57 Si, 18.02 Cr, 11.87 Ni, 2.16 Mo, 0.80 Cb. Each of these percentages was multiplied by the "potency factor" indicated for the element in question along the axes of the diagram, in order to determine the chromium equivalent and the nickel equivalent. When these were plotted, as point X, the constitution of the weld was indicated as austenite plus from 0 to 5% ferrite; magnetic analysis of the actual sample revealed an average ferrite content of 2%.

For austenite plus-ferrite structures, the diagram predicts the percentage ferrite within 4% for the following stainless steels: 308, 309, 309 Cb, 310, 312, 316, 317, 318 (316 Cb), and 347.

Dashed line is the martensite/M + F boundary modicifation by Eberhard Leinhos, "Mechanische Eigenschaften und Gefugeausbildung von mit *Chrom und Nickel legiertem Schweissgut*," VEB Deutscher Verlag fur Grundstoffindustrie, Leipzig, 1966.

Figure 10

The Schaeffler Diagram for Stainless Steel Weldments

(Schaeffler, 1949)

indices were the carbon, nickel, and chromium equivalents. For stainless steel welding, it was also necessary to draw a diagram indicating the interaction between the nickel and chromium equivalents and the failure mode to be avoided. Calculation of the carbon equivalent (for low-alloy and plain carbon steels) or the nickel and chromium equivalents and Schaeffler diagram location (for stainless steels) will provide useful measures for choosing weldable materials from a set of otherwise equal candidates. This does not, of course, guarantee that no other welding problems will arise. As more detail is generated, additional concerns will rise. These simple figures of merit will, however, put the material choice on the right footing to serve as the basis for these later decisions, with two major problems limited or eliminated.

4.4. Casting

Elegantly simple in concept, casting can be, if considered in detail, a complex process. The simple idea of pouring molten metal into a mold, obtaining a component which has nearly the net shape desired, is the basis for a large fraction of the tonnage output by the world's heavy industries. Yet the problem of obtaining high-quality castings can be quite complicated.

In terms of material selection, the casting problem is analogous to welding. Nearly every metal is castable, the only question is whether that material is prone to difficulties which raise the cost of the casting produced.

Two simple measures of castability are the melting point and the molten range of the alloy. The melting point is the temperature at which the alloy is all liquid metal. The molten range is the temperature change required to go from the liquidus (all liquid metal) to the solidus (all solid metal). Cast irons tend to have carbon contents between 4% and 7% because a composition near the eutectic allows the lowest possible melting point and a narrow range of temperatures for solidification. Steels are more expensive to cast because their melting point is higher, and their solidification range is wide which results in difficulties with mold filling due to inconsistent flow of the metal.

There are a number of specific limits on castability of particular metals. Some, like tin, oxidize so readily that it is difficult to maintain purity in the molten metal. Other metals have unfavorable thermal properties in the liquid, limiting the tendency to fill a mold. Exceptions like these are important, and because they deal with particular details they are more amenable to treatment through a narrative, rather than an analytical, knowledge representation.

The most important information about casting during material selection is recognition that all metals have certain limits on their mechanical properties as a result of the as-cast microstructure. In particular the effects of large grains and directional grains must be considered. Large grains in general lead to poorer fracture resistance. Directional grains, generated by dendrite growth from the mold surface inward, both reduce fracture resistance and contribute to anisotropy in the properties of the component obtained. Thus a cast component will have generally poorer strength and dramatically poorer toughness. The first property is reflected in tables of as-cast mechanical properties; the second is not necessarily obvious.

Another important issue in casting is segregation during solidification. One may consult a phase diagram and see that while passing through the solid plus liquid region the composition of the solid formed changes as the casting cools. This results in segregation. Not only are the grains formed in the cast structure large and usually directional, they are also of varying composition. In the foundry, trace additives are used to improve grain fineness and limit segregation.

Liquid metal will be considerably more reactive than the solid. For that reason, particularly if an alloy is being considered for a casting, it is necessary to evaluate the reactivity of the individual metals which comprise the alloy. As an example, the author worked on a system for rapid-solidification processing of a silicon alloy which is used as a grain-refining agent in the iron casting industry. The alloy is used to improve grain fineness and limit segregation as mentioned above. To measure the cooling of the metal in order to determine the amount of supplemental cooling required to be supplied to the system, a thermocouple, sheathed in stainless steel, was employed. The sheathed thermocouple was rated for temperatures well above the range to be measured. However, the thermocouple failed before data could be taken, not because of the temperature, but rather because the stainless steel is soluble in liquid silicon.

The use of casting, like any process requiring closed molds, also dictates certain geometry issues. It is necessary to provide a draft angle in order to remove the part from the mold; therefore the as-cast part will not have parallel sides. In addition, a plane (not necessarily a flat one) which defines the separation between mold sections must be identified. These geometric limits impose restrictions on the shape of the part to be cast, and are comparable to limits on geometry of forged parts.

Casting is a manufacturability issue which is not readily handled through analytic means. In general terms casting is not as much of a problem as other manufacturing processes because there are established classes of materials - cast iron or brass, for instance - which are routinely used for castings. Requirements for cast part are flexible enough that adjustments at the foundry to allow a more castable material are common. If, however, casting of a nontraditional material is being contemplated, the rules given above about melting point, solidification range, and reactivity provide a basis for reasoning about the castability of any metal.

4.5. Geometry

Although the focus of this chapter has been on material selection, some useful and general rules about geometry have emerged from this research. In reviewing the literature on forging, welding, and casting design, a strong analogy to stress concentration developed. This provides useful guidance to an engineer who may not be familiar with the particulars of forging or casting design, but anticipates one of these processes may be used to produce the part being designed.

A few simple rules apply to all of these situations. To a large extent they represent engineering common sense, yet they also represent information that is typically treated at the detailed level, late in the design. These rules are equally useful during earlier stages, when the geometry is just beginning to be developed. Using these rules may not result in an optimally manufacturable part, but they will result in a part which is amenable to optimization with minimal changes.

- Use generous radii through shape changes. The reason this is important from a strength viewpoint is a lower stress gradient results from minimizing the geometric gradient. The same analogy used to explain stress concentration, force flow, provides the explanation of this effect in molding processes. Rather than force flowing through the material, the material itself must flow through the mold or die to fill. Sharp corners lead to problems with closely spaced stress isobars; they also lead to flowline constrictions in the material filling the mold. For machining, small-radius tools are in general more expensive than larger radii.
- Avoid sudden, or sharp changes in section thickness. This is related to the first rule. Section changes lead to stress concentrations, unless they are quite gentle. Similarly, section changes lead to mold and die filling problems.
- Small features are more expensive than large features of the same geometry. Small features involve greater wear on forging dies, on casting patterns, and on cutting tools.

4.6. Summary

To choose a material with good manufacturability, it is necessary to recognize the potential for certain problems which can make a manufacturing process more expensive. In general, all four of the branches of manufacturability discussed dealt with similar considerations. Each of the issues is treated, at present, as a detailed problem. The author has demonstrated simple, more abstract methods which provide economical anticipation of manufacturability during material selection. The author concluded in the first chapter there is a need for economical and predictive heuristics, useful for design. This chapter contains demonstrations of some of the forms these heuristics can take, and indicates methods for developing them.

Machinability may be assessed by the use of an index which can be calculated based on readily-available mechanical properties of the metal. Ductility and strength are the key parameters, with thermal conductivity playing a role as well. Ductile metals are more difficult to machine than brittle because of the energy absorbed dues to plastic deformation prior to the chip breaking away. Higher-strength materials are more difficult to machine because of the greater energy required to exceed the breaking strength and remove metal.

For the purposes of evaluating formability during material selection, it is necessary to consider two indices: the flow stress, an indicator of the energy required to form the part; and the strain-hardening exponent, a measure of the limits on deformation. The strain-hardening exponent may be estimated based on three mechanical properties which are readily available: the yield strength, the ultimate strength, and the reduction in area. For the purposes of preliminary decisions about material choices, one can neglect temperature effects and still obtain reasonable comparisons of formability.

Weldability of aluminums can be anticipated quite simply from the general alloy class or from the temper designation. Anticipating weldability of steels during materials selection requires consideration of indices which are based on alloy chemical composition. The indices are the carbon, nickel, and chromium equivalents. For stainless steel welding, it is also necessary to draw a diagram indicating the interaction between the nickel and chromium equivalents and the failure mode to be avoided. Calculation of the carbon equivalent (for low-alloy and plain carbon steels) or the nickel and chromium equivalents and Schaeffler diagram location (for stainless steels) will provide useful measures for choosing weldable materials from a set of otherwise equal candidates.

Two measures of castability are the melting point and the molten range of the alloy. However, evaluation of castability requires consideration of not only whether the metal will fill the mold and take on the correct shape, but also what mechanical properties the cast part will have. Further processing, such as heat treatment to improve properties and finish machining, must be considered.

The indices and figure-of-merit comparisons discussed here are necessarily estimates. Index values may be used for comparison, but the values must be interpreted. In particular, index values for the candidate materials cannot be taken simply as a rank ordering system. If the values are spaced some distance apart on the scale, then such ranking is not unreasonable. But if the values fall near one another, order becomes meaningless and one must simply conclude the materials are nearly equivalent under the parameter. During material selection, either of these two conclusions is useful evidence.

It is worth noting that manufacturability knowledge for materials selection, especially that for forming and welding, deals with anticipating and preventing problems more than with simply ranking of performance in the production process. This makes the handling of manufacturability quite similar in flavor to the handling of failure prevention. That is one reason that both of these problems are addressed in this thesis.

For early design decisions, detailed issues are not of general concern, although some details can be important from the beginning. Early, pre-detail decisions are not concerned with optimizing, because in that uncertain and ill-defined situation optimization has no real meaning. Rather, the need is to make decisions which account for major problems, and are robust to accommodate later refinement.

It is important to note the mix of knowledge representations which were useful in examining the question of materials selection for manufacturability. Data and numerical evidence was not developed for direct use in the materials selection decision, but information of this form did serve an important purpose when considering the reasonableness of the approximations made.

While major emphasis in this chapter was placed on analytical information, there were also situations where narrative information was the more appropriate form. In the next chapter, the author will explore more fully the narrative form of knowledge representation and its application to design.

Chapter 5
Material Selection for Failure Prevention

This chapter is a report on the author's research into anticipating and preventing failures during material selection. Improper selection of materials is the root cause of a substantial fraction of the failures in the literature. Typically, the improper selection deals with a material characteristic which was not evaluated during material selection.

Performance, as noted in the second chapter, is concerned with the system serving its intended functional purpose. Quality is concerned with preserving performance consistently throughout the service life, thus with preventing unanticipated loss of function. Since a failure is the source of loss of function, preventing failures is an important ingredient to improving product quality during design.

More effective design for failure prevention is a subset of more effective design for product quality. This chapter is concerned with methods for detecting and eliminating possible sources of failure in the design, early in the design process.

Predicting the occurrence and effects of failures in the product is the primary purpose of reliability analysis. Reliability, in turn, is one of the most tangible aspects of product quality. But evaluating a completed design's reliability is not the same as designing for reliability.

An approach to failure prevention is needed which will catch and prevent those few details which might cause problems. This approach should meet the requirements for design knowledge and design tools outlined in the first chapter: economy in the thinking required to apply the knowledge and tools, coherent connections to both knowledge already in use by design engineers and the information available to designers about the product design, and flexibility in the ways they can be used.

The customary approach to detecting and correcting errors is during the process known as checking, that is, some time after the error has been made and only shortly before the design is produced. Approaches to design checking will be considered. Another approach to failure prevention is to develop predictive information about failures which can be used to anticipate and correct potential problems as a natural part of design decisions. A predictive approach, and an example of how it can be applied, will be presented, followed by a general discussion of how to implement design for failure prevention methods. All of the approaches to design for failure prevention discussed here can be used by the unassisted engineer, and can also be effective when supported by a computer.

5.1. Design Checking

It is generally true that designs contain errors. The errors may not be significant, and in fact may not be detected for a long time after the designers have finished, but there are usually errors present. Although looking back through the design, checking it, is expensive, it is even more expensive to correct the design when the error is first detected as a malfunctioning product. Therefore effort is put into looking for errors before the design is committed to hardware.

Diagnosing a failure which has taken place in a machine is a challenging problem. Recognizing a potential failure in an operating machine is at best inexact, relying heavily on intangibles such as experience and insight. Anticipating such problems in design is still more difficult. The main requirements of the design can be evaluated, and these performance requirements generally point to a small number of failure modes to be avoided. But there are a large number of other failures of concern which have not been made explicit by the performance requirements, leaving to the engineer the problem of searching for potential failures which may be hidden in the design.

Complicating the question of searching economically for failure modes is the cost of correcting problems. The greater the time from commission to correction of an error, the greater the cost of correction (Halpern, 1978). More importantly, early errors are also more costly because they tend to span a greater portion of the program, affecting a larger number of details and enlarging the scope of the correction. Early detection of problems leads to less costly corrections, implying a need to search for problems and errors as early as possible. Late searches may be more effective because the greater wealth of detail aids the search, but they also lead to more expensive corrections.

Ideally, the approach to failure detection should enhance the engineer's ability to recognize errors as they are made, detecting and eliminating the ingredients of a failure before they are put in the design. Correcting errors right after they are made minimizes the cost of correction. Errors and other sources of problems should be detected during the course of making a particular decision, and corrections made before the decision is completed. Immediate correction of potential problems is predictive error prevention, actively anticipating problems rather than passively reacting to them when they manifest themselves.

One can generalize the kinds of errors to be sought and corrected into two classes: explicit errors and implicit errors. An explicit error involves a mismatch, either between two components or between a component and the operating environment. The most obvious example is a physical mis-fit between two components, as in features not aligning or an interference between a shaft and the hole to receive it. The mismatch might also deal more with function than fit, such as a mechanism member undersized or poorly balanced for the load and speed at which it is intended to operate. In general it can be said that explicit errors can be checked on a part-by-part basis, working from the detail drawings. Explicit errors lie in areas that are normally evaluated as the design is developed.

Implicit errors involve problems which are not necessarily considered by the engineer developing the design. An implicit physical misfit, for instance, would be a hole and shaft which will assemble by the nominal dimensions, but which can have an interference at the maximum material conditions of the tolerances. Another implicit error is a material which is poorly suited to the service environment. In general, implicit errors cannot necessarily be trapped by checking individual drawings.

It is useful to think of reliability in terms of a basic method of reliability analysis, the fault tree. If engineers can build complete sets of fault trees for a system, then they have completed failure mode identification. The question is, how does the engineer know the set of fault trees is complete? Finding root causes is a difficult aspect of human problem-solving, and in psychology experiments, humans have been shown to be ineffective at generating the ideas required to build or complete fault trees. Gettys, *et. al*. (1980) asked subjects to list all the possible causes for a generic problem, such as a car failing to start. Both lay persons and experts on the problem areas produced remarkably incomplete lists of causes or courses of action. Even if incompleteness in itself were not a problem, subjects have also exhibited great confidence the incomplete fault tree was complete. Fischhoff, Slovic, and Lichtenstein (1978) gave incomplete fault trees for similar problems to their subjects, asking them to assign probabilities to each branch shown listed. Regardless of what was missing from an experimental fault tree, which the experimenters described as going so far as to be "obviously incomplete", subjects consistently assigned low probability to the branch labeled "all other faults". Reviews of failures examined in the literature, such as in "Failure Prevention and Prevention" (1986) leads one to conclude the same effect is present in engineering: failures are not generally caused by inadequately evaluating potential problems, they are caused most often by problems which were not considered. Designers seem to be confident they had considered every important branch of the system's fault tree, when at least one additional failure was operative in the design. An example illustrating this situation will be discussed in detail later in this chapter.

Failures occur only rarely as the result of a factor which was not evaluated correctly, instead their cause usually lies in a problem which was not considered at all. Put another way, failures are caused by details that fall through the cracks. Improved reliability (hence improved product quality) is obtained by searching a design for cracks and inspecting the details the cracks contain. Simply detecting the potential for a failure places the engineer in a position to prevent it.

When undertaking evaluations of potential modes of failure, there is a need to consider a list of potential failure sources which is as complete as possible. But evaluating the severity of any one of these modes can be expensive and evaluating all of them is impossible under even the most generous budget. Thus the engineer wishes to make a minimal, economic evaluation of modes of failure, and to evaluate only those few which truly are issues. The first step is to decide which problems require consideration. Almost all failures are preventable, and many failures can be prevented without extensive analysis. Only those failure which cannot be eliminated entirely require evaluation to determine their severity.

Complementary to the two kinds of errors, explicit and implicit, checking a design for errors has two flavors. Drawing checking, exhaustive checking of notes, shapes and dimensions on each part drawing, is standard procedure for most projects. Design checking has a broader meaning, and addresses questions of whether the complete system will meet the needs of the customer. Design checking is not undertaken regularly.

Drawing checking is very effective in dealing with explicit errors. Dealing as it does with the detail drawings, drawing checking focuses on the smallest details of the design and considers them exhaustively, one by one. Exhaustive drawing checking is feasible because there are a finite number of details to be checked. A typical engineering drawing might have fifty checkable details per sheet; a typical, large-scale project could run to 1000 sheets of drawings. 50,000 details may seem like a large number, but at only a few minutes per detail, even a very large project can be checked completely for explicit errors in a few thousand man-hours. These estimates reflect the author's recent experience in a large mechanical system. Smaller projects are checked in proportionately less time.

Most engineering organizations have design checking systems in place, many of them highly structured and formalized. It is commonly assumed the best way to find errors is to be very methodical, on the assumption that method will lead to thoroughness. The author has visited companies where a formal design checking system is nominally in place, but most of the productive error trapping seems to be accomplished by showing the design to experienced engineers. These experienced engineers are not methodical, but they are thorough and effective. The method seems to be simply remembering the right detail at the right time.

The broader problem of checking in its other flavor, which the author calls design checking, deals primarily with implicit errors. There are a variety of reasons one cannot simply conduct an exhaustive search for implicit errors. The most obvious one is that there are a large number of unique, potential failures to consider. Hundreds of material failure modes are known, and some are not well-defined. Material failures represent only one of several classes of ways the design can fail. Unsafe operator decisions, occupational safety, and the general problem of product liability are other kinds of failure which may be dealt with in the design, not considered in this chapter, which should also be considered by the engineer.

Checking a design for failure modes can be considered a search problem. There are many different flavors of search, thus different methods of search support are required. Winograd and Flores (1986), for example, make a distinction between library cataloging systems which allow the user to obtain further information for a known author and title (as a librarian might wish to) and a system which supports searches based on subject areas (as a researcher might wish). The difference is between the questions, "how can I find this book?" and "what is relevant as I approach this subject area?" But neither of these questions is comparable to the problem of searching for unidentified problems in the design. A reasonable library search analogy to design checking is the problem of maintenance and conservation: "which of the books in the library requires attention to remain in good condition?" No reasonably complete solution can be obtained by searching the library's catalogs. Although inspection of each book seems to be the sensible solution, brute force search of each volume in the stacks is not a feasible option either. Rather, a majority of the books selected for conservation measures are simply noticed during handling by the circulation staff or by users. Some volumes which need repair are not noticed and go back to circulation.

The search for potential failures also cannot be undertaken by brute force alone. Direct, exhaustive search depends on evaluating every known mode of failure in turn, but the purpose of the search is only to identify those modes which may require evaluation. An exhaustive search based on evaluating individual failure modes is uneconomical, and probably infeasible for most designs.

Considerably less effort is required for a search based on heuristics rather than on exhaustive search methods. A heuristic is any method whose value is not universal, its usefulness governed by the particular problem it is applied to (Koen, 1985). Heuristic searches are useful in a number of situations where procedural searches may be too difficult, too time-consuming, or simply impossible (Newell, 1969). However, a heuristic search, in contrast to a fixed procedure, does not guarantee success.

Rather than concentrate on details, it can be more economical to rely on general aspects of the design - the component types, the intended service environment - and to use those clues to identify situations similar to the design where failures did arise. From this identification, the engineer is in a position to take action against failures which may arise. Having identified a failure mode of interest, economical consideration of reliability is achieved by considering whether that mode is preventable - or already prevented - in the design. An assessment in this manner does not provide a direct measure of reliability, but it does provide the means for the engineer to accomplish two important tasks which enhance product reliability:

- eliminate many potential problems soon after they are introduced; and
- compile a list of failure modes which require analysis, with confidence the list is reasonably complete.

Accomplishing the first task allows the design to be made more reliable by preventing unanticipated problems and limits the costs of later corrections to the design. The result will be a higher-quality design. The second task provides the justification for the expense of detailed analysis where it is necessary with minimal investment in evaluations which are not needed. Together, these two improvements make maximal use of existing knowledge for failure prevention with a reasonably small investment in time and cost required to carry out design checking.

A very useful additional feature would be for the act of checking to improve the design judgement of the engineer, so that the accrual of checking experience would produce progressively fewer errors requiring correction. There is some concern that automation of design tasks tends to reduce the opportunity for engineers to develop their own expertise (Cooley, 1986); certainly a design checking tool, of all possible design tools, should have the opposite effect, and contribute to improving the engineer's judgement. The most successful methods discussed in this chapter also provide for improving the skill of the user. This skill improvement develops an engineer's predictive knowledge about failure prevention, reducing dependence on feedback from checking or component failures and reducing the cost of error correction.

Searching for a potential failure might seem to be related to failure analysis, which takes place after the failure has occurred. While it is true that much of the useful information available for design checking has been generated through study of prior failure analyses, the methodology of failure analysis is not well suited to design checking. There is a substantial difference between failure prevention and failure analysis. Although the two deal with the same body of knowledge about modes of failure, the uses of the knowledge, hence the useful forms the knowledge can take, are quite different in character. In particular, a failure analyst knows that a failure has taken place, and has a clear idea of the location of that failure. The failure itself provides clues about its origins. The failure preventer, in contrast, has no direct indications of what failures might take place, nor of where they might occur.

There are a number of examples illustrating the relationship between failure analysis and failure prevention. The semisubmersible oil rig "Alexander Kielland" capsized and sank in the North Sea, caused by a fatigue crack originating at what was thought a minor structural change which had not been evaluated. Easterling's (1983) failure analysis of the Kielland was based almost entirely on data, such as fracture surface photographs and microhardness traverses through weld cross-sections, that was not available to the designer; in fact it would not have been available to anyone prior to the failure without destructive testing. The failure of the skywalk in the Hyatt Regency of Kansas City was caused by ambiguity over who was to design a structural connection (Petroski, 1986); in the end it appears no one did (Marshall *et. al.*, 1982). The clues used to determine the causes of the failures were taken from the wreckage, after the failure. The clues required to have prevented these failures were, arguably, available, but the failure analysis could not have been carried out had the location and mode of failure been unknown. While in failure analysis, the engineer must find a cause from the clues about a particular detail of the system, in failure prevention the engineer must find potential causes, from almost no clues, about every detail of the system.

5.2. Abstract Modes of Failure for Design Checking

Every failure has certain key ingredients which can be generalized and linked through logical operations. This knowledge can be used to develop abstract, logical models of the failure - a Material Failure Logic Model or MFLM (Marriott and Miller, 1982). In this section MFLM's for several failure modes will be discussed, then applied to design checking.

An MFLM can be expressed as a fault tree, showing the necessary ingredients for a given type of failure. Only those ingredients which are observable in advance of the failure are recorded, since evidence available only after the failure has no predictive value. The MFLM can be applied to design checking by searching for the components of the failure in the details of the design and fabrication process. If all of the ingredients are present, then the failure is logically feasible in the design. This information provides guidance on failure prevention to the designer. Having identified a mode of failure, the required ingredients for the failure are evident, and prevention requires eliminating one or more of the contributing causes. Marriott and Miller (1982) and Handrock (1984) provide a number of MFLM's for failures common in the literature of failure investigations.

Strain-age embrittlement provides a useful example. The constituents of the MFLM for strain-age embrittlement are summarized in Figure 11. Susceptible metals are the rimmed steels; killed steels do not exhibit this failure mode. As the name implies, the failure requires straining and aging to achieve the embrittlement which leads to failure. If the steel is strained more than about 4%, then it is sensitized. Aging can be either moderate temperatures over a long period, as in service in a warm location in a plant, or a higher temperature for a short term, as in elevated-temperature service, repeated steam cleaning, or perhaps a heat treatment. It is possible for post-weld heat treatments, ordinarily applied only to the weld itself, to provide the aging adjacent to the heat-treated region. After aging, the steel is brittle and can be a location with significantly poorer fatigue strength, or a fracture initiation site.

To consider how well an abstraction helps in design checking, the reader might like to use the information just given on strain-age embrittlement to answer the following question: to what extent is strain-age embrittlement a concern for the construction of pressure vessels? All the necessary rimmed steel = susceptible material

+ cold work over 4% strain

= sensitized for embrittlement

+ age: ambient (long time) or elevated (short time)

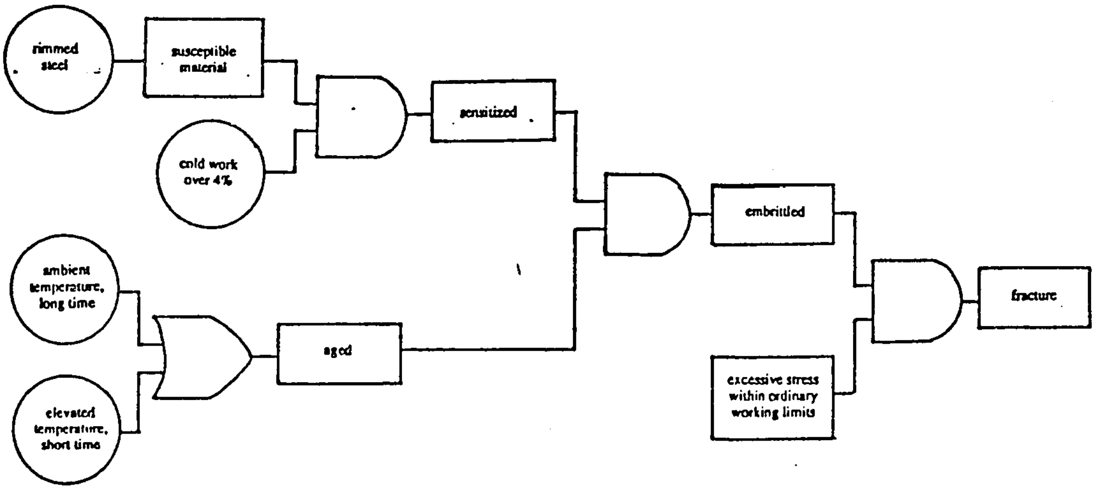
= embrittled

+ stress exceeding (unknown) limit

= fracture

as a material failure logic model

(after Handrock, 1986)



as a fault tree

Figure 11

Strain Age Embrittlement

information, in abstract form, has been provided. The answer to this question will be discussed in a later section of this chapter.

It is not necessary to understand the microstructural mechanism for strain-age embrittlement to understand its prevention. All that is necessary is to locate the contributing, macro-scale ingredients and prevent at least one of them which is connected through a logical conjunction ("AND") within the fault tree. Thus strain-age embrittlement can be avoided through selection of a killed steel, by limiting plastic deformation during fabrication, or by avoiding the potential for aging. Another approach is to heat treat the entire component after cold forming, overaging and eliminating the embrittling effect. Failure prevention can be achieved by anticipating and eliminating a contributing cause of a problem in advance, or by recognizing an unavoidable problem and adding a correcting change to the design.

Untempered martensite is another embrittlement mechanism which can lead to fracture in the absence of an initiating defect. Susceptible steels include any which can form martensite. A surprisingly large range of manufacturing steps can bring about the cycle of heating and rapid cooling necessary to form a local patch of martensite. Examples include spot grinding, such as is done to remove the parting line "flash" from a forging or the sprue from a casting; electric-discharge machining (EDM); arcing during electroplating; electroetching part identifications; or welding if not planned and executed correctly. Ironically, arcing during improperly conducted magnetic particle inspection has also led to the problem ("Failure Analysis and Prevention", 1986).

Any effect which heats a small location of the steel above the temperature required to form austenite provides the potential for martensite formation on cooling. Cooling is usually rapid in a local heating situation because of the quench effect provided by the bulk thermal mass of the adjacent metal. Some of the famous Liberty Ship brittle fractures initiated at welding arc strikes. The behavior of the fractures suggest untempered martensite could have been a contributor to the failure. Arc strikes are a common source, but not the only source. For instance, the cable drum on a mine shaft elevator failed because of brittle patches of untempered martensite. The heating was provided by friction heating when the wire rope occasionally slipped, locally, on the drum; cooling provided by the metal adjacent to the slip.

Preventing martensite formation is difficult because the only ingredients involved are a susceptible material (nearly all steels, to a greater or lesser extent) and a cycle of heating followed by rapid cooling caused by any of a wide range of manufacturing steps. Bednarz (1988) found untempered martensite to be at the root of roughly ten percent of a sample of over 600 failure cases reported in the literature.

Untempered martensite is an example of a problem which is not easily prevented. An approach to prevention during design is to recognize the potential for the problem, and anticipate that additional care (thus additional cost) will be necessary during manufacturing to assure the process is properly controlled. A more certain (and more expensive) approach is to add tempering to assure that, if formed, martensite does not remain in its most brittle state.

Handrock (1984) provides some useful intermediate ideas which simplify application of the MFLM methods. The simplest concept, used above in the discussion of strain-age embrittlement, is a susceptible material. Some materials can experience a particular failure mode, others can't. For strain-age embrittlement the susceptible material concept provides a clue to the potential for a problem. At the point a rimmed or semi-killed steel is selected (which can be inferred from the amounts of silicon and aluminum specified in the chemistry) the engineer can recognize a susceptible material. A susceptible material is one capable of exhibiting a particular mode of failure.

Sensitization is a step closer than susceptibility to a potential failure. A material is sensitized if part of the failure mode has progressed, but the process can be reversed without damage. A rimmed steel strained over four percent has been sensitized. A strained steel which has been aged is even more sensitized, but the brittle material condition can still be removed through appropriate heat treatment. Untempered martensite due to an arc strike is a sensitized state. Sensitization indicates ingredients of the mode are accumulating toward completing the requirements for failure. The damage mechanism is preventable, however, if the potential for the failure is detected. In a sensitized material, progress toward failure is fully reversible.

Sensitization need not apply only to physical changes. The idea can be extended to describe the design itself. Many components are cold-formed from sheet or plate stock due to the higher energy cost of hot-forming. Thus choice of a rimmed steel for a cold-formed component always leads to a design which is sensitized to strain-age embrittlement. Untempered martensite provides another useful example. Many steels are susceptible, so no useful decision can be made on that basis. But some manufacturing steps, such as welding, electric-discharge machining, or grinding, provide a significant risk of the necessary cycle of heating and rapid cooling. Here the possibility is so great, the design can be called sensitized because of the significance of the combination of material and process. Sensitization toward a mode of failure provides a strong warning to the designer before the design is committed to metal.

The stepwise progression provided by use of the concepts of susceptible and sensitized materials reflects the flexibility needed by a designer. During material selection, the engineer requires reminders of problems to which the materials being considered may be susceptible. Later, as more detailed decisions are made, cues based on sensitization can provide more specific information. Finally, after all the detailing is complete, the full MFLM can be applied to reason about the nature of each potential problem identified during design checking. But to use only the detailed aspects of knowledge about material failures is to lose most of the value of the MFLM approach.

A majority of the useful working experience in application of MFLMs has been through manual use, where it has been successful. A useful lesson can be drawn from the attempt to apply this method of reasoning in a computer-supported system, where the concept was applied in conjunction with an exhaustive search approach.

5.3. Exhaustive Search for Design Checking.

The MFLM method has been applied in FERRET, a computer-based system described by Lue (1986), which demonstrates some of the limits of exhaustive searches for design problems. The emphasis in Lue's work was to locate potential failures without excessive identification of problems which were not a genuine concern, that is, to minimize false positive results. In his application, both the software and the data store grew quite large to deal with a limited number of failure modes. Extending the system to handle even a moderate fraction of the MFLMs available would have expanded the data store to an unreasonable size. Not only the design, but also a fully planned manufacturing process are required as inputs for the program; both are required to be stated in considerable detail. FERRET is only useful during the last stages of production planning, or for diagnosing problems during production because it deals with the details of both the design and the manufacturing processes and facility. In the author's opinion, FERRET is used too late in the product development process to be considered an aid to design.

Because of the need to account so thoroughly for the details, the developers of FERRET focused on how to represent those details, rather than on the problem of identifying problems for engineering review. FERRET provides an object lesson for design tool developers. It is necessary to focus on the design problem, regardless of how hard it is to deal with that problem. It is easy to slip to back to a focus on what the computer is capable of doing, rather than on what engineers need to have the computer do for them.

The primary disadvantage of the FERRET system is that it imposed a single logical structure, the MFLM implemented to deal with the details of the design and manufacturing plan, on all problems. Many failures are not necessarily easy to handle in this structure. Coombs (1986) notes that building such models forces relations which had been ambiguous or implicit to either be made explicit or be ignored. These relations had not been explicit any earlier because their structure was not clear. The structure imposed by FERRET may or may not reflect the structure actually in the relation. The actual structure depends on context, and selection in advance of a single structuring approach severely limits the use of the tool containing that structure. FERRET could be used for those cases which fits its intended use, but it cannot be expanded to become a general-purpose tool for design checking. This limit is not necessarily a problem, except that the distinction between failures which can be modeled effectively and those which cannot remains unclear. It appears that only hard experience will allow the distinction to be made, which does not seem to be a good investment of time and effort. There is also a substantial amount of time and effort required to obtain and understand a FERRET report, and incomplete results do not seem worth the investment.

The heavy dependence on details - both on the details of the design and on the details of the failure modes - is a limit on the usefulness of the FERRET system. Design checking by exhaustive search is time-consuming: putting a description of the design and the intended manufacturing steps into the right format, maintaining descriptions in detail of the manufacturing processes, and maintaining descriptions of the physical and functional connections between the individual process locations in the plants. It is important to note, however, that this limit is a fault of the implementation, not of the MFLM concept, which does have some value as discussed in the previous section.

It was noted earlier in this chapter that drawing checking by exhaustive search is possible, in fact is a standard way of doing business. Drawing checking by exhaustive search is feasible because the number of details to be checked are finite, even if large in number, and because the kinds of errors to be trapped - mainly, errors leading to mismatched geometry - are few. Thus explicit errors can be managed through the use of exhaustive search, and contemporary experience implies it is the preferred method.

There are an infinite number of details and their combinations in a given design. It is not economical to start with the design's particulars and search each possible permutation of them for the ingredients of a problem. The result of an unsuccessful attempt to develop an approach on a detailed basis - FERRET - has been described in this section. There are a finite number of failure modes, but there are enough of them that they are virtually infinite, because evaluating every one is prohibitively expensive. Design checking for implicit errors is not as conveniently handled through exhaustive search, as indicated by the experiment with the FERRET checking system. There are more kinds of errors, and the method for locating them is not necessarily as clear as the methods for trapping explicit errors. However, there are approaches to design checking for implicit errors which can be reasonably successful which will be discussed below.

5.4. A Failure Case Study

In 1973, a horizontal pressure vessel at a fertilizer plant in Potchefstroom, South Africa, was being filled with anhydrous ammonia from a railroad tank car. With the contents at about 0.6 MPa (90 psig) pressure and about 5 degrees Celsius temperature, a large section of one of the ends failed by brittle fracture and was projected a short distance away. The ammonia spill resulted in 22 deaths, most of whom were residents of houses nearby. The description which follows of the events is largely derived from the account given by Campbell (1979).

The design conditions specified to the builder in 1967 were 1.8 MPa (250 psig) pressure and 50 degrees C temperature. The vessel had been built of semi-killed carbon steel. The end had been cold-dished to form the main radius, then hot-formed on a smaller radius to meet the cylindrical body, a common practice in the pressure vessel industry. Two plates had been welded together prior to forming the end, since a single plate was not wide enough for the diameter required.

In 1971, when a hydrostatic test was required by law, the owner applied for and was granted postponement of that test, since it would interfere with production. Instead, extensive ultrasound inspection of the vessel was undertaken, which revealed large areas of delamination in the dished end. The inspection also indicated some lack of fusion in the weld used to join the two plates which formed the end.

Despite the fact that the lack-of-fusion defects were products of the original manufacture and had not caused any problems, they were prohibited by British Standard 1515, the code governing this particular vessel. The defects were ground out and repair welded on-site. One repair weld failed to pass inspection, and was ground out and repaired again. The reweld had hydrogen embrittlement problems and cracked, making necessary a third repair weld.

The code did not require postweld stress-relief heat treatment for vessels of this size, since the wall thickness was below one-half inch. Accordingly, no heat treatments were applied.

After repairs, the vessel was hydrostatically tested at 2.4 MPa (360 psig) in late 1971. After repairs to a valve in 1973, the vessel was pressurized again at 2.4 MPa (360 psig) to test the valves. This took place six weeks before the failure. Both tests were at ambient temperature.

After the failure, there was considerable interest in this vessel since at the time there was some uncertainty about the propensity of ammonia for causing stress-corrosion cracking in steels. Despite considerable investigation of the stress-corrosion cracking question, no positive evidence of stress-corrosion cracking was found.

No initiating defect was ever identified in the failure. The fracture surface and measurements made after the failure indicated the metal was so brittle that any defect, no matter how small, could have been the initiator. Although no single initiation site for the fracture was located, all of the most likely candidates sites were adjacent to the weld seam on the dished end, in the heat-affected zone. It appears the failure was due primarily to embrittlement of the steel through strain-aging. Part of the strains were caused by cold forming the dished end, and part appear to have been contributed by the residual stresses in the weld heat-affected zone. Because the cracks fell along the heat-affected zone, it can be inferred the heat input from the repeated repair welds contributed to the aging process.

This case has a remarkably wide range of potential uses in design. The most obvious application deals with the warning about strain-age embrittlement. Cold forming dished ends on vessels is common, as is use of semi-killed steels, which cost less than killed steels. At low levels of silicon and aluminum in the chemistry (the elements used to deoxidize or kill the steel), a semi-killed steel can exhibit the properties of a rimmed steel.

As the name implies, semi-killed steels are a compromise between the lower cost of a rimmed steel and the superior properties, principally fine grain size, of a killed steel. Strain-age embrittlement is found in rimmed steels which have been cold worked beyond about four per cent strain. After that, only aging - tens of hours at 100 degrees C or hundreds of hours at slightly above ambient - is required to obtain the brittle condition. Figure 11 summarizes the required ingredients for this failure.

Despite the fact that stress-corrosion cracking does not appear to have been the cause of this failure, the case does remind the engineer of that issue. There are some cases where ammonia can lead to such problems.

An important lesson is the effects of the defects originally located by ultrasonic inspection. The lamellar tears, which appeared to have been caused by the cold forming operation, had no influence on the final failure. The lack of fusion weld defects did not appear to have impaired successful operation for several years prior to the repairs. One would not expect these plate discontinuities to have affected fitness for purpose because of the absence of significant cyclic loadings. However, the repairs led to series of additional problems. There is little doubt that the repair welds contributed both additional strains and the elevated temperature required to age the material to a brittle state. The cold cracks (indicative of hydrogen embrittlement) generated by the repair welds were considerably more dangerous than the original lamellar tears. Lack of fusion in a factory weld, which led to the repairs, is also instructive as it is generally assumed factory welds will be defect-free.

The case leads one to question what influence an overpressure "proof test" may have on continuing structural integrity. There have been components which failed in proof testing, so a single test on delivery may at times be useful. One wonders, however, whether the proof test is valid when conducted under other than operating temperature and filling conditions. The author has witnessed proof tests of several pressure vessels, typically air receivers, and all were treated in a fairly casual manner by plant personnel, despite the fact the vessel is taken to roughly twice its normal pressure. There have been cases where damage generated by a proof test led to failures in service.

Comments on repair welds and proof tests are mentioned to demonstrate that there are useful avenues of digression available away from the main point of a case study.

The design specifications provide a particularly useful lesson. The conditions specified to the fabricator were only the maxima - maximum pressure and temperature - on the assumption these constituted the worst loading conditions. The highest service temperature is important in a design limited by its ductile strength. But for brittle fracture considerations, the lowest service temperature is far more important. The transition temperature for most steels falls between the design temperature specified by the client for this vessel and the service temperature for liquefied-gas service. Rather than assuming design loads can be defined by a single point in the service conditions, it is more appropriate for the client to ask for the vessel to suit a range of conditions, leaving the designer to define the envelope of limits on the design. In the case of a pressure vessel, the designer is obliged to ask about this point. Had the vessel been specified for liquefied-gas service, implying low temperatures, it probably would have been made from a fully killed steel, eliminating entirely the potential for the failure which took place. It is not always recognized that anhydrous ammonia is a material subject to cryogenic conditions. This case provides a good example of the value of flexibility, rather than formality, in the information provided to the designer.

In an earlier section of this chapter, an abstract description of strain-age embrittlement was given, along with a question about its prevalence in pressure vessels. The abstraction alone did not provide a good basis from which to answer the question. This case, however, provides clues which allow the reader to express an opinion on locations in a typical vessel where strain-age embrittlement might be found, types of service where it is especially of concern, and how it might be exacerbated during the service life. The case study is easier to follow, comprehend, and remember, making it a richer source of information.

The Potchefstroom case is by no means unique in its ability to convey a wide range of useful messages. Most cases, filled with details and plot twists like a detective story, can provide insight into a wide range of problems, depending on the reader's viewpoint and goals at the time. A case study achieves flexibility and economy in knowledge representation because the narrative is capable to encompassing a wide range of potential problems, and different searches with different intents may each yield something useful from a given case.

5.5. Case Studies for Design Checking

Material failures may be described through examples as well as through the physical and chemical principles which define their origins and behavior. A simple method of searching for failures, then, is searching the design and a descriptions of cases where failures occurred, comparing them for similar features.

There is substantial evidence in favor of case studies as a useful learning method as well as for engineering evaluation. St. Denis (1989) observes, "empiricism, which is the conclusions drawn from a study of past failures, has the golden quality that it guides one away from making gross blunders." Turkstra (1986) mentions effective use of case studies in management and law as he argues for greater use of case studies in engineering education. Peters and Waterman (1984) rely extensively on case studies to make a book on effective management comprehensible to almost anyone. Students in psychology and medicine routinely review cases as part of their education. Hunter (1989) discusses several failures to draw lessons in general about preventable failures.

Dreyfus and Dreyfus (1986) argue expert knowledge takes the form of case studies. A standard approach to interviewing experts for expert system development is to use the context of specific problems to obtain inference rules for the knowledge base. Frequently, when an expert discusses a problem with a client, the evaluation and solution will be couched in terms of "another situation just like this". Although the similarity may not be visible to the client, the solution obtained through reasoning by analogy to past experience generally is effective in solving the problem. A popular expression at the author's present employer puts it this way: "good judgement comes from experience, and experience comes from bad judgement."

Petroski (1985) makes an important point about case studies in engineering. A failure case is always useful, because there is always at least one specific point about the failure from which a lesson can be learned.

A successful design, however, always carries with it the uncertainty that a more economical solution might have been obtained by less conservative decisions. For avoiding failures, it is clear that more information is available from a failure case than from a success.

There have been failures - the space shuttle Challenger (Rogers Commission, 1986) and the DC-10 cargo door (Eddy, Potter, and Page, 1976) - where the potential for failure was known, but not dealt with due to management intervention to meet cost and time limits. The disadvantage the engineers in these cases faced is their evidence was not convincing to managers. Clear identification of a similar failure in a related design, and its consequences, provides a basis which managers - technical or nontechnical - need in order to be able to consider the magnitude of the problem. Because failure mode identification can be undertaken early and economically, budget and schedule for evaluation can be better anticipated, limiting late problems which can lead to adverse management intervention.

Using a case study is a reasoning process which Hofstadter (1979) calls prototyping. The specific example of a single case provides a prototype of an entire class of similar cases. Psychologists refer to the representativeness heuristic (Kahneman and Tversky, 1972), through which properties of a population are inferred from a single sample. This heuristic is one of several methods which are used to simplifying reasoning about complex problems. People relying on this heuristic can err by attending to irrelevant features, or by disregarding relevant features of the problem. Error is minimized in the application of case studies to design discussed here because relevance is the primary focus of the engineer referring to the case.

In its simplest form, a design tool based on case study review could be any of the published case history collections. A half-step superior to that would be individual companies with specific problems maintaining a "corporate experience" file which would survive independent of personnel changes. There is also the obvious step of relying on a computer to assist in storage and review.

The case study approach can be applied at any point in the design process. Case studies provide for flexibility in the design approach by allowing the individual engineer to choose the time to use the method. Use of case studies for early feedback limits early commitments by helping the engineer avoid potential problems without the need to pin down details of the situation. Because the matching between case and design is made on a qualitative basis, through broad similarities, it is not limited to use at only the end of the design process. The review of case studies provides the design engineer with concrete examples on which to base reasoning about both the nature of the problem and its severity. The engineer can also become better at locating and preventing failure modes on his own, based on insight gained during review of cases.

The case study approach achieves the goals of economy, flexibility, and coherence. A full analysis of a failure mode is not required for the engineer to become aware of the need to give a possible problem area more thought. As mentioned above, failures rarely result from a problem which has been incorrectly evaluated after being identified. A logical model of a failure, such as the MFLM already considered, allows the engineer to move directly from identifying a problem through a case study to reasoning about prevention. Thus simply identifying a potential failure mode, which can be achieved through the use of case study comparisons, is a substantial step toward prevention.

A case study can be applied to identify potential problems at any point during the design process. The time invested in the search can be dictated by the time available at that point in the project. The results, which take the form of cases identified as having some relation to the project, can form the basis of quick and effective decisions on how to proceed. The relevance of a particular case can be determined after a quick review. Some problems can be corrected without much evaluation beyond noticing the potential problem. Evaluations which do become necessary can be prioritized based on the full list of potential problems and the time and budget available at that point in the design process. If evaluation exceeding available budget is required, the project manager can be provided with evidence on which to base the decision for additional investment.

Design checking by reference to case studies of past failures succeeds because most errors are repeated. Wearne (1979) noted that, of a dozen major disasters (coal mine fires, offshore platform capsizes), not a single accident was caused by unknown phenomena; all could have been prevented by use of existing information. The problem is locating the information which applies to the situation the engineer faces and presenting it in a manner which leads to economic and effective action by the designer.

Design checking by reference to case studies relies on the intelligence of the engineer, rather than attempting to provide intelligence in the design tool. The method supports the engineer in locating potential failure modes, an area where support to enhance human skills is appropriate. Deciding on what to do with these modes, once identified, is an activity at which engineers are generally quite effective. Supporting failure prediction in design, such as by search of case studies, is therefore an area in which development of tools to support design is appropriate and needed.

Case studies as a basis for design checking offers substantial aid in failure mode identification, and does so in an economical manner. No effort is wasted on evaluation until the engineer has made decisions regarding which of the identified failure modes are of the greatest concern. In the preceding section, the amount of useful information which can be found in a single case study was demonstrated.

5.6. Predictive Failure Prevention

The well-established methodology of reliability analysis (see, for example, Henley and Kumamoto, 1981) is best suited to the question of evaluating the potential reliability of a completed design, applying such methods as fault trees and various statistical measures to the evaluation. In some cases, such as in electronics design, it is possible to learn about design improvements from the reliability assessment. Designs which can be readily handled through traditional, statistical reliability are large networks made up from limited component types. In the case of electronics, the components are semiconductors, resistors, conductors, and capacitors; in process plants they are valves, pipes, fittings, and vessels. In these large-network situations it is possible to learn from past experience and cost-effective component tests, relying on statistical evidence which applies because of the large numbers of individual, readily characterized, components involved. But when the components grow in size or are uncommon, there is a decreasing amount of valid evidence on reliability, and consequently an increasing uncertainty about what the overall reliability of the system might be. Uncertainty over systems with larger or unusual components applies equally to electrical and mechanical systems.

In evaluating the reliability of an electronic circuit, the focus is on the probability one of the large number of components will fail and lead to loss of function in the system. Individual component failure statistics and the interactions between components are well-characterized because the logic of a failure's effects can be mapped directly from the circuit diagram. Thus statistical methods are appropriate for predicting the aggregate behavior of a sizable population of similar individuals which make up the system. Individual components which have the greatest influence on unreliability may be identified and replaced in the design with more reliable ones.

It is more difficult to characterize mechanical system failures because the modes of failure are so varied, because the interactions between components are not as clear, and because a mechanical system is generally made of a smaller number of components, drawn from a much wider variety of types. The details of the service environment, mode of loading, and manufacturing and service histories for a material all must be considered in order to predict the likelihood of a failure. For only a few, limited classes of component, such as rolling-element bearings, statistical characterization is possible due to the clearly-defined manufacture and service (Juvinall, 1983). Even in these cases, some aspects - loss of lubrication, dust or water damage, overloads or shock loads - can lead to component behavior significantly different from the probabilistic prediction. Bearing manufacturers classify severe loadings as outside intended service conditions, although most offer guidance on down-rating to suit loading conditions outside the norm. There are designs for which "improper" use of a bearing is unavoidable. As a result, designing for mechanical reliability must focus on modes of failure, that is the origins of component failures, rather than the effects of failures. In industry, reliability assessments of mechanical systems deals primarily with estimating, after the fact, what the reliability of the system might be.

It is not necessary to consider reliability improvement from the basis of finding a numerical measure of the probability of a successful system. All that is necessary to improve reliability is recognition and prevention of potential failures. In general the knowledge needed to prevent a failure is considerably more compact than that for evaluating it. To improve a component's low-cycle fatigue performance, an engineer must address improving material properties and reducing the effects of component geometry and the load history. Despite the value of the damage accumulation theories and what is known about fatigue microcrack formation in predicting the fatigue life for a given component, an ability to perform fatigue analyses contributes only subtly to understanding how fatigue performance could be improved. Failure mode identification, in addition to being required to accomplish a reliability assessment, leads the engineer directly to think about failure mode prevention. Recognizing the potential for a problem is a substantial portion of the cure.

The foundation of reliability evaluation is failure mode identification. Building a fault tree to describe the origin and consequences of a single failure is a tractable problem, and given a list of potential problems there are tools and techniques which can assist in constructing the fault tree and producing the relevant analysis. But without correctly anticipating all potential problems, fault trees will not be complete, leaving potentially important modes of failure out of the reliability evaluation. More importantly, failing to identify a failure which has potential in the design means the engineer will not consider its prevention. Support is not readily available, in the form of either methods or direct tools, for assistance during failure mode identification during design.

In the past, mechanical designs have been built relying on a series of prototype tests to provide the feedback. Economics, both time and budgetary limits, now dictate that these tests "in iron" be minimized. Comer (1987) interviewed an engine designer at a heavy equipment company, who explained that his group has a goal of building only one prototype, as proof of concept, immediately prior to beginning production and sales. The group is disappointed if a prototype test does not produce results closely matching analytical and numerical predictions. Without feedback from physical tests there is a much greater need for thorough assessment of potential failures in the design. Engineers no longer have the luxury of fixing the design after fixing the prototype. The strategy for finding problems can easily become an exhaustive search which, of necessity, remains incomplete. Within a constrained problem area, such as engine design, experience is important for keeping design decisions effective and for recognizing and eliminating potential problems. The discussion will return shortly to the role of experience.

There is a serious economic limit on the process of failure mode identification. Because failures seem to arise from the interactions between isolated details, it is nearly impossible to identify every possible failure. Knoll (1986) points out design checking is often limited by the cost of errorhunting rather than by a rule defining the stopping point of the search based on the possibility of remaining, unidentified problems. Since exhaustive search is not really feasible, the choice of stop rule is a major factor in the successful detection of potential failures in the design. A goal for developing suitable methods for searching is to identify as many potential problems as possible with minimal investment. Achieving this goal allows a majority of time and of the analysis budget to be expended on the possible failures of greatest importance.

In the past, judgement based on experience has played a major role in locating failures. Judgement about failure prevention was simply one of the creativity problems left to the engineer, relying on what Knoll (1986) calls "a rather fuzzy consensus of a number of experienced people". This approach was reasonably effective because engineers had extensive experience with design, were involved in the production process, and had worked for years with particular types of problem.

Experience is the best basis on which to check designs and correct potential problems. This conclusion was unanimous among the engineers with whom the author discussed design checking. The most effective approach to improving judgement is through experience. Yet, because experience is "fuzzy", it has not been possible until recently to explain the reasons for its importance. With the advent of expert systems, there has been renewed interest in the psychological principles of eliciting an expert's experience, and a clearer explanation of the value of experience has been developed.

Effective engineering experience is not limited to analytical or technical matters. An individual's entire experience of a prior event comes into play in the memory - not only the evaluation of it, but also what was seen and what emotions were felt in reaction. Evanson (1988) points out that when an expert is interviewed about a problem, a great deal of the useful information is prefaced by expressions like "I see", or "I have a feeling". He warns this subjective, observational knowledge is probably even more important than the direct analytical information.

If experience is the best basis for failure mode identification, then the most useful supplement to human experience is the one which most resembles a real experience. An abstract description of a failure mode is only a portion of the complete experience of a component failure. Being abstract, an analytical evaluation lacks many of the ties - visual and emotional, for instance - which helps them become lasting memories for use in future problems. In other words, by evaluating a particular failure mode only through use of abstractions, such as mathematics or statistics or even the logic of an MFLM, the engineer develops skill in making the manipulations required to carry out the evaluation but develops little skills in recognizing and preventing such a failure in future assignments. Further, the use of abstractions presupposes a structure for what is known, and the user is forced to spend time meeting that structural requirement before the technique can be used.

A narrative case study encompasses a broader view, coming closer to an approximation of genuine experience. Since the goal is simply to identify modes of failure which may require evaluation, reviewing experiences with similar components, similar materials, or similar manufacturing processes can help the user to locate likely problems. Since two contexts are provided, the case study and the design being reviewed, and since the narrative provides some limited visual and emotional content as well, the user can learn by using the case study, and may be able to avoid a problem in the next design through anticipation, rather than reacting after the fact.

A review of case studies as a design checking method does have one drawback: it is still reactive in the sense that the engineer makes a series of decisions, and then goes back to review them. An additional step to make error-free design more achievable would help the engineer to consider case studies and other failure information while decisions are being made. This requires the engineer to be knowledgeable about failures and their prevention, which comes mostly from experience. It is possible to help the engineer accumulate experience faster, using mnemonic techniques which make the information easier to recall.

5.7. Mnemonics to Aid in Predictive Failure Prevention

In many situations a failure may be prevented by having a single piece of information available. If this is true for every possible failure, then there are a reasonable number of these pieces which, if remembered when needed, can be used by the designer to anticipate and correct problems during design, instead of afterwards during design checking (or even later, such as in service).

In some cases, a single cue can remind the experienced engineer of a large volume of warnings. Simple statements which provide warnings about potential problems are important gems of wisdom which require capture. In this research, such statements are called one-liners, reflecting their compactness.

The contrast between this compact, "one-liner" form and the ways technical knowledge is usually conveyed is dramatic. A recent, well-publicized problem provides a useful example.

Creep is a mode of failure which occurs in all metals, in fact in almost all materials, dependent primarily on temperature. In metals, creep will take place when the service temperature exceeds about half the melting point. (Creep being a phenomenon driven by the thermodynamics of the solid, temperatures must be considered on an absolute scale. This is also a useful one-liner.) One might assume that creep is found only in elevated-temperature service, and therefore the need to consider creep in design would be obvious. A problem with a fixed disk drive for personal computers (Willet, 1988), however, indicates there are cases where creep is a problem at room temperatures. A flag indicating the location of the zero sector on the disk platter, essential for continuing correct operation, was attached to the disk shaft through an interference fit, achieved by use of a screw which applied circumferential stresses to the collar holding the flag. This design would have been an adequate solution, except for the fact that in production the collar was die-cast from a zinc alloy. Zinc's melting point is low enough that it can creep at room temperature. Substitution of the zinc die-casting in production for the aluminum sand casting in the prototype led to the problem. The cost reduction associated with that substitution was dwarfed by the warranty claims and loss of sales generated by the problem.

Avoiding this problem required recognition of the potential for room temperature creep in the selected material. In narrative form, one only needs the observation that zinc creeps at room temperature. This one-liner could have prevented a very expensive problem. A more general, and therefore more flexible, one-liner is the observation that all metals creep when raised above roughly half their melting point on an absolute scale. In the more abstract version, the engineer must still stop to evaluate the melting point of the metal in question with respect to room temperature.

The disk drive example illustrates the principle of economy in knowledge representation. A single sentence would have been sufficient to prevent a severe loss of quality on the disk drive, reducing costs to the supplier, the vendor, and computer owners. The Zinc Institute (cited in Willet, 1988) does in fact warn creep can take place, but the warning is spread through several paragraphs in their information brochure. Succinct presentation of the facts involved is vital to failure mode identification and failure prevention.

The pressure vessel case reviewed above indicates another useful one-liner: Repair welds can generate as many problems as they solve. There are any number of experienced engineers, including the author, who have been forced to learn this from a series of repairs.

One-liners are most effective when they serve as a point of entry to recall of larger history. A case study dealing with unstable austenite illustrates the point. The conveyor in a heat treating furnace for steel parts is heated above the austenite transformation temperature, thus ordinarily steels are not appropriate conveyor parts. Two typical solutions, both expensive, are to use nickel alloy parts or to keep the conveyor parts cooled, such as with a water spray. One conveyor was designed to reduce these costs by using a Hadfield steel for the conveyor chain. Hadfield steels are made austenitic at room temperature through the addition of a large amount of manganese. The designer's reasoning was that an austenitic steel would not be affected by heating because the austenite transformation is avoided. After a series of heating-cooling cycles as the conveyor ran, however, the chain links failed because the manganese had precipitated as carbides, allowing the steel to become ferritic at room temperature and brittle because of the precipitates. In this example, the designer relied on "stable" austenite to solve a problem, when in the service environment the austenite wasn't stable.

The mnemonic derived from the Hadfield steel case points out the importance of a context: stable austenite isn't necessarily stable. In absolute terms, stabilized austenite simply transforms back to ferrite so slowly that we ordinarily can assume it is stable. There are a number of mechanisms which can cause problems, and loss of assumed austenitic properties. These include formation of carbides or alternative phases (through any of several mechanisms), embrittlement, stress-corrosion cracking, as well as plastic deformation. The one-liner provides the engineer with just enough pessimism to use austenitic steels thoughtfully, but the connection between the one-liner and the greater supply of knowledge about austenite and its instabilities must be familiar if the engineer is to have any real benefit. One-liners do not represent a complete statement of failure prevention knowledge, rather they provide compact, mnemonic entry points into more complete descriptions of potential problems.

One-liners are economical because they provide only the pointer to a potential problem. They are flexible and easy to use because they are simple in format. They are reasonably mnemonic, so the engineer has a better chance of using these concepts during design decisions, rather than to make corrections after the design decision has been made.

One might imagine there are an unreasonably large number of one-liners which might apply to a given project. The primary function of the one-liner is to improve access to what engineers already have in their experience. It was argued earlier in this chapter that complete experiences, not simplified abstractions, form the memories which are recallable for application to solve new problems. The complete memory of a case may not, however, be easily recalled. The one-liner provides the memory cue to help recall of more complete information about a problem. Engineers cannot be expected to simply memorize a large number of one-liners out of context. The appropriate technique is for one-liners to be presented as an adjunct to a case study or other failure-prevention knowledge; and for engineers to develop their own one-liners about their own experiences.

5.8. Design Applications

The success of a design checking tool can be measured by its ability to prevent errors which had, in the past, been prevalent. This section will review the success of some failure-prevention methods, then outline the form and uses of a somewhat more general tool which could be built on the ideas developed in this chapter.

Organizing a book of cases for the purpose of providing an error-prevention resource could require extensive indexing and cross-referencing. The organization of such a volume becomes fixed at publication, while engineering needs continue to change. A computer-based approach can be arranged to make each search without an index by looking through the text of the cases for matches with keywords. Rather than continually updating an index, the computer can search rapidly enough with no index at all. It is also important, however, to recognize that an exclusively computer-based system has limits. Some of these limits will be considered in the discussion of two experiments with this kind of design tool.

Nitta *et. al*. (1980) discussed an information system, called HIRIS-DR, which was intended to share failure experience among the 20 divisions of Hitachi, Ltd., in Japan. The system operates by retrieving microfiche copies of failure reports and design checklists. Selection from the fiche supply is automatic, based on keywords supplied by the user. Use of the system generated an immediate improvement in productivity. After three years, failures in one division were reduced by about 85%, and design reviews were carried out in one-third or one-quarter of the time required prior to introduction of this system. The particular feature of systems like HIRIS-DR is they record and circulate corporate experience. Two divisions which are otherwise isolated from each other can share useful information.

In a project connected to the research reported here, Bednarz (1988) used commercially available text retrieval software on a personal computer to assemble a simple system to search for relevant cases. It became almost instantly useful to engineers who were asked to test it. Although no data on industrial experience has been collected, an experiment was conducted with mechanical engineering undergraduates taking a materials course. For a power plant boiler tube, an unfamiliar component, students were able to identify and begin reasoning about preventing, on average, about half of the potential failures of interest. Some students did considerably better. For a gear, with which students were generally more familiar, they were more effective, identifying about three-quarters of the potential failure modes of interest. From these experiments with inexperienced students, it is possible to imagine how effectively it might be used by experienced engineers. Even undergraduates could anticipate and move toward preventing half of the potential failures.

Students identified failures which were not related to the assignment, which Bednarz called "misses". A miss is simply a potential problem obtained during a search which is not related to the design being reviewed. The engineer conducting the review is interested in relevance and severity of potential problems, and a reasonable number of misses should be acceptable.

In part, misses are the fault of the keywords used to conduct the search. Bednarz reported that searches in his system were limited by the (relatively) small number of keywords in the text of each case. Other words could be used to name the same concept - such as the synonymous terms cold cracking, hydrogen assisted cracking, and hydrogen embrittlement - so recall was limited not only by availability of a concept within the case studies, but also by the user and case study sharing the same term for that concept. He supplied limited sets of synonyms to help the user find more productive keywords, which did help somewhat. An improvement would be a system which carried a large set of associations to synonymous keywords, allowing a wider-ranging search for associated cases, bringing with it the risk of a larger number of misses.

A more fundamental aspect of misses is characteristic of text retrieval problems, the relationship between recall and precision (Blair and Maron, 1985). Recall refers to the number of relevant references obtained in the search, as a fraction of the number of relevant cases actually available. Precision is the number of relevant references as a fraction of the total references retrieved. The misses obtained by the undergraduates in Bednarz' work are examples of precision less than unity. Precision and recall have a relationship much like strength and ductility in metals; to make improvements in one it is usually necessary to sacrifice part of the other.

The limiting relationship between recall and precision does not pose a problem for a design review application. Recall is improved by the fact that different cases have features in common, generating redundant supplies of information. The most common failure modes are also the most represented in a collection of failure case studies. Misses are not a problem since skipping over a case study of no interest requires little effort. Reviewing a case study summary which is of no interest also requires little effort. Misses are not expensive, and therefore lower precision - more misses - can be tolerated in order to achieve greater recall.

Notable in both Bednarz' and Nitta's demonstrations is the absence of complicated assumptions about the user's needs. The systems are fast and effective in delivering to the engineer descriptions of situations that should be avoided. The engineer must then decide what further action on a given mode may be appropriate.

The effectiveness of the basic concept of narrative design checking systems have been shown by Nitta *et. al*. (1980) and by Bednarz (1988). An earlier section, discussing the FERRET application provides a contrast, a more exhaustive approach which was not as successful. In particular the difference in computing resources required is of interest: the less successful method required all of the resources of a Unix-based engineering workstation; the more successful method was demonstrated by Bednarz on an ordinary personal computer.

The search for possible problems can be conducted from several directions. Some problems are associated primarily with certain materials, others with particular fabrication processes or component types or service conditions. Some failures require combinations of these. As each of these new conditions is developed through design decisions, the search for failure modes can be repeated for more detailed development of the problem.

Component names and types are logical entry points into the link of associations which retrieve cases for error trapping in mechanical design. Because of similarities in loading, geometry, service environment, and other features, components of similar type experience similar failure modes. Associations to similar case studies may be used to advantage as a simple feature of CAE work packages currently being marketed for mechanical, concept development design, such as the ConceptStation (Aries, 1987) the Mechanical Advantage (Cognition, 1987), or I-DEAS (System Dynamics Research Corporation, 1990).

The addition of a warning retrieval subsystem to any CAE system could operate in the same manner. The engineer supplies pertinent information on a query screen to the design checking module. Information given for an early search could be as vague as component name or type, intended material, service environment, and manner of loading. In some cases, the search would have to be selected as the primary process. On more advanced systems, the pattern-matching process could run as a time-share in the background while the engineer works on the drawing and calculations. When the search is complete, the module would indicate that cases and related information are available for review.

There is some potential for incorporating the MFLM concept into a design checking system. The MFLM can be used as the basis for reasoning about prevention once a failure mode has been found relevant to the design. At early stages, with little detail yet defined, the search on a broad basis would yield cases from prior experience, warnings about material susceptibilities, or perhaps sensitization. A principal application the author sees for the MFLM is in failure elimination after detection by other means. Having identified a failure which may be a problem in the design, the engineer can use the MFLM to reason about alternatives for prevention. The MFLM does provide economical and flexible knowledge about eliminating the potential problem.

It is not assumed that all potential problems can be identified at an early point, nor that all the problems identified will require correction. It is up to the engineer to make these decisions. The purpose of a design review subsystem is to provide early cues, reminding the engineer of possible problems. Evaluation of a potential problem may or may not involve calculations; in many cases the logic of the failure ingredients are more directly applicable to preventing incidence of the failure. Further searching on more detailed information can be supported in the same manner, either from keywords provided by the engineer or by automatic action based on information accumulating in the design file as further details are developed. Having identified a failure mode which is of concern, the engineer can use the design checking tool to obtain information about corrective action and prevention. Other tools can be used to carry out whatever analysis is required, at the direction of the engineer.

Most CAE packages for mechanical engineering store a complete problem context in the design file. The results of searches for problems should be incorporated into the long-term memory about that component. When a similar component is to be designed at a later date, efficiency indicates the earlier design file will be modified as necessary to develop the new design. Experience can be recalled as well as the part drawing. There is a strong interest in using corporate memory available through computer-aided design files to avoid "reinventing the wheel" by storing drawings for later adaptation to new uses (Gaertner, Nieva, and Argoff, 1986). Retaining a connected description of what problems were considered helps to avoid reinventing the square wheel - that is, later engineers know not only what workable solution was obtained in the end, they also have available a list of some of the pitfalls which had previously been avoided so the engineer adapting the design can avoid bringing problems back into the design.

A simple extension of the design checking subsystem described could include a "tell me about" digression feature. During the search process, upon reading a case study which seems to indicate a potential problem, the engineer can turn away from the review of cases to find more information about that specific problem. At that point the MFLM for the failure can be presented, along with examples and references. This is, in essence, what engineers want when they turn to a reference book to read about a failure mode.

This "tell me about" diversion is the basic operating idea of some relatively new software products based on the hypertext concept (Conklin, 1987). A useful model of hypertext software is a set of index cards connected by threads, which represent cross-references. The original idea was to allow researchers to accumulate notes about their work without constraining them to the linear, sequential organization forced by notebooks and conventional documents. At present, hypertext does not seem well-suited to a flexible case study retrieval system because the links themselves are a primary aspect of the data structure. Typically, these links must be created when a new notecard is added to the data store, whereas in the hypothetical design checking tool discussed above such links should be created by the act of searching (Frisse, 1988). Still, the basic idea may be a useful one if a suitable reorganization of priorities can be achieved.

The computer-based tools being used to support design have not, until recently, provided a basis of support for evaluations other than those quantitative and numeric in nature. Current developments lead in the direction of tools which can support qualitative reasoning as well - such as review of related failure case studies - particularly when the users do a lot of the reasoning for themselves. Further development in the same direction, however, will be necessary before the hypothetical tool described in this section can be realized.

5.9. Summary

This chapter contained a report on a few experiments in tools to aid the engineer in handling a difficult problem: failure prevention. A successful approach to design for failure prevention is review of failure cases in related designs. The relation might be similar materials, similar function required of the component, similar manufacturing steps, or similar service environment and conditions, among others as well as combinations.

Three approaches to design checking have been discussed and contrasted. The first, the FERRET system, was quite structured in its approach, and exhaustive in its executions, and for those reasons more restricted in its usefulness. The MFLM approach used as a basis for FERRET is not at fault; the basic approach has been used successfully in other implementations and has been shown above to have useful applications. The structuring imposed by the architecture of the FERRET system limited the system's usefulness. The system itself was designed to make decisions, limiting the system's ability to support design except in specific, narrow situations.

The second system considered was relatively free of structuring assumptions on the part of the developer, and provides simple information about failures in similar designs to the engineer, to be the basis of decisions about failure modes which require consideration. Several different methods may be used to implement a case study review system, including computer-supported systems as well as less complex techniques. Having less structure, a case study system is more flexible, allowing the system to be adapted to a wider range of applications. Coherence is achieved by the context carried with the case study narrative. Decision-making is left to the user, so the decision can be influenced by the full context of the problem under review, allowing simplifying assumptions to be made, and permitting adaptive interpretations of the results.

Tools based on case study reviews are relatively simple to use, allowing the engineer to remain in control of the process of deciding what to do about the problem after it has been located. In fact, the engineer is required to make decisions about failure modes, enhancing skill development and reducing long-term dependence on the tool. The case study method discussed in this chapter has the capability to provide support exactly where it is needed, in the problem of recognizing subtle patterns which may indicate a problem.

Case studies may be used for both general and specific aspects of corporate memory. Cases may be reviewed indiscriminately in a search for related problems which may guide thinking in other designs. A complete design file can include cross-references to cases which were helpful during the engineering process, allowing a later engineer, adapting a design to new uses, to avoid problems which had originally been dealt with.

The lesson from the comparison of the FERRET system with case study approaches is that, by selecting a particular detailed structure as the basis for solving a particular problem, the creator of a design tool necessarily imposes limits on the ways the tool can be used, limiting flexibility. This is a general conclusion with validity in a wide variety of situations, and provides an important guidance point for design of design tools.

One-liners are a technique which engineers could use to improve their ability to recognize possible flaws as they are first introduced. One-liners internalize mnemonics which serve as compact cues to larger, more complex information which need not be retained in memory at all. One-liners create a situation conducive to predictive failure prevention at the earliest possible moment, instead of during design checking.

Designing for failure prevention shifts emphasis from finding and correcting errors and problems, to anticipating problems and making the original design error-free. Effective recognition of potential problems is the key issue in both feedback, design checking methods and feed-forward, predictive design for failure prevention. To achieve the forward-looking approach to error-free design, the tools and methods used for design checking must provide an avenue moving away from the more common feedback approach.

Chapter 6
Knowledge Engineering for Design

In the two previous chapters, examples of the knowledge necessary for engineering design have been presented. Borrowing the concept of "knowledge engineering" from expert systems developers, the author will in this chapter develop from those examples some conclusions about the knowledge which engineers use for design, how it is represented, and how it is structured. The concepts of flexibility and coherence will receive a more detailed definition, and a method will be presented for evaluating how well a knowledge structure fits the needs of a design engineer.

6.1. Knowledge Representation for Design

The issues of manufacturability and failure prevention, discussed in the fourth and fifth chapters, illustrate the variety inherent in the knowledge necessary for effective design decisions. In this section an organization of the useful forms knowledge can take will be presented, defining them in the context of their level of detail and the form of reasoning necessary to apply that knowledge in design. Based on the organization described in this section, the section to follow contains expanded definitions of flexibility and coherence, and describes means for achieving those desirable properties when recording knowledge for design use.

It is first necessary to discuss the concept of knowledge representation. Like design, knowledge representation can be a verb as well as a noun. A knowledge representation may be an object, the structure in which knowledge is recorded. Knowledge representation is also the act of creating and using such a structure. Because of the constraints on computer-based reasoning, knowledge representation in artificial intelligence research refers to particular structures for recording knowledge in a form amenable to manipulation by a computer. Perhaps the most well-known knowledge structure in artificial intelligence is the production rule (Hayes-Roth, Waterman, and Lenat, 1983), which imposes the format of predicate logic on the knowledge it records. Rather than discuss particular knowledge representation schemes, this chapter will focus on determining how well a representation meets the needs of a designer.

Within the definition of knowledge representation for computer purposes there lies a root definition which can be applied regardless of the intended user of the knowledge. Knowledge representation is the collection of structures which put the knowledge into a form convenient for the intended user to solve problems. Two features of this definition are important: the required form is driven by the needs of the user, and the knowledge is being collected because it will be used as a basis for problem-solving.

No product can succeed if it was developed without considering how it will be applied by its users. The same is true for new ideas in engineering. The intended product of engineering research is new ideas to be applied, during design, to development of new products. Therefore ideas which are intended to be applied by design engineers must be presented in a way suitable for use during design. Since most research products are not of a suitable form, it is necessary to add an additional step, after the generation of new information, to restructure the information for use. Nadler (1981) calls this step reduction to practice. However, the use of the word "reduction" implies there is some lessening of value in that step, one not worthy of a researcher's efforts. The author prefers to apply the concept of fitness for purpose to new information. Researchers must ask themselves whether the information they have generated is fit for the purpose of supporting new product development.

The reason for conducting engineering research is to advance the product development process. Therefore the reason for spreading information produced by research is to expand the design engineer's resources. The main support required of engineering research is improved methods for solving problems.

The attribute which characterizes a knowledge base is the degree to which the information it contains is interconnected. Two separate knowledge bases cannot simply be combined to form a single knowledge base because they have no connections between them; even when combined, the two knowledge bases will remain primarily independent. This limit appears to be the underlying reason why two separate expert systems cannot be combined to solve still-larger problems. However, more than one knowledge structure can be incorporated within a single knowledge base. To improve connections between separate knowledge bases, two changes are required: use of a larger variety of knowledge structures within each knowledge base, and development of a foundation of connections between knowledge bases.

Expressing all of engineering within a single knowledge base is a desirable goal which may, eventually, be reached. At present, knowledge for engineering design is splintered. Each discipline within engineering has its own researchers, and there is very little work being done to deal with the interactions between disciplinary areas. Since the author contends that design deals primarily with the interactions between areas, the previous sentence can be restated by saying that very little of contemporary engineering research supports the design function, not because the targets of research are selected improperly, but because the form in which the results are distributed are not suitably fit for the purpose of solving design problems because they are insular in nature and lack the necessary connections to the rest of engineering.

Expert system development involves long and intense efforts by domain experts, as well as the knowledge engineer. Winograd and Flores (1986) observe that in most cases the collection and structuring of knowledge created by expert system development is as significant a research contribution as the expert system itself. This is an indicator that, in addition to knowledge-producing research in engineering, there is also a need for research, such as the research reported in this thesis, into organizing and representing knowledge for design use.

There is a substantial gulf to be spanned between the ways engineering information is generated and the ways it is used. While this research has not produced a universal method for making the connection, concepts which can serve as guidance for thinking about the problem have been introduced. The concepts are the attributes which knowledge for design must have: flexibility and coherence. These two concepts were introduced and informally defined in the first chapter; the section following will expand on the definitions and demonstrate how they can be applied in thinking about structuring knowledge for engineering design.

6.2. Definitions of Flexibility and Coherence

The primary need of an engineering designer is the support required for economical problem-solving. Two working principles were introduced in the first chapter: flexibility and coherence in the supports for problem-solving. This section contains more complete and explicit definitions of these two terms. Before those definitions are presented, it will be useful to consider the relationships between the different knowledge forms presented in the fourth and fifth chapters.

Figure 12 indicates two axes which can be used together to clarify the relations between different knowledge forms used in engineering design: the nature of the information, and the level of detail involved. Information can be used for reasoning which varies in form from quantitative to qualitative. Quantitative reasoning deals with things which can be stated analytically or numerically, such as equations or data. Qualitative reasoning handles more descriptive information, such as a narrative description of a process, or phrase descriptions such as the individual elements of a Material Failure Logic Model.

The level of detail can vary from the particular to the general. Examples of particular expressions are data, such as the machinability data used in the fourth chapter, or descriptions of single events or processes. General information is more abstract, for example an equation or a Material Failure Logic Model.

Although the two axes shown in Figure 12 are continua, it is a convenient approximation to identify entries along an axis by the axis endpoint. This divides the knowledge representation map into four quadrants, which are also shown in Figure 12. The examples discussed above in the definitions of the

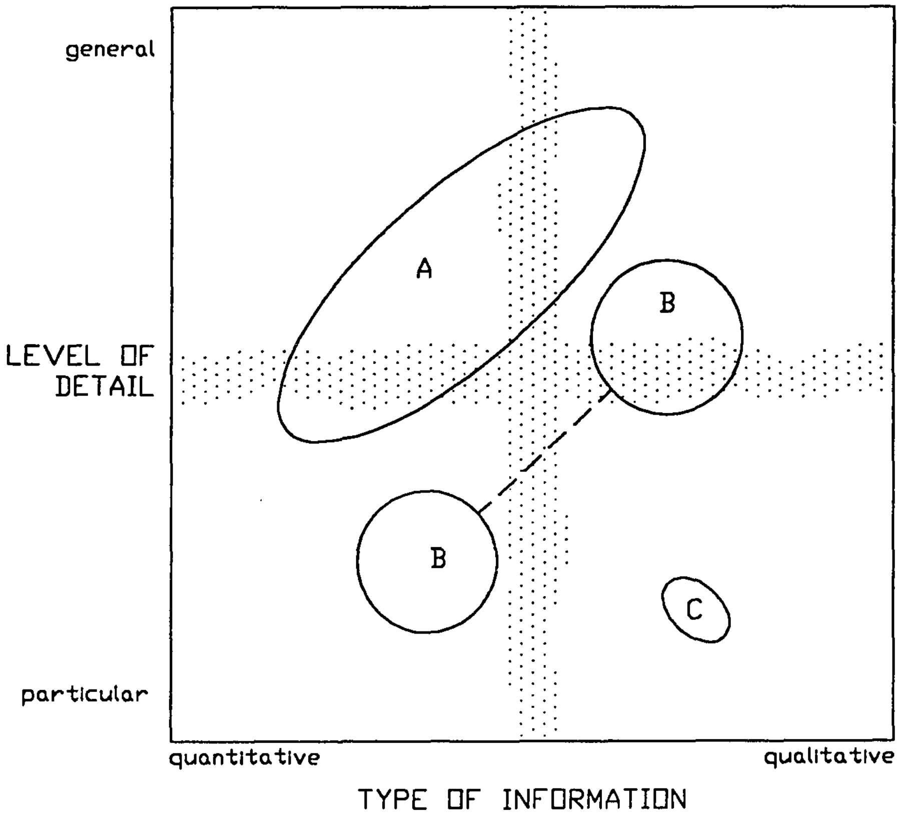


Figure 12

A Knowledge Representation Map

axes appear in the appropriate quadrant on the map. The simplified, quadrant version of the knowledge representation map is useful because it eliminates the need to make a quantitative judgement of a location on an axis.

The view of the knowledge representation map shown in Figure 12 is the view the author will use most in the discussions to follow. However, there are other dimensions to the knowledge representation map. The third axis, not shown in the two-dimensional view of Figure 12, is an axis which shows the continuum of knowledge. In practice, this can be an axis labeled with the names of the various disciplines within engineering. Another view of the knowledge representation map, which appears later in this chapter, will make use of the third axis.

The knowledge representation map is used in this chapter to reason about the knowledge used by engineers for problem-solving during design. It was created by the author as a tool to support an assessment of the state of the engineering knowledge base. There are other potential uses of the knowledge representation map which are not indicated in this chapter, such as to evaluate tools intended to support design decision-making.

The knowledge representation map can be used to refine the definitions of flexibility and coherence. By refining the definitions at this point, it is possible to clarify the needs of knowledge representation for design.

Because the fundamental problem being addressed is allowing the engineer to include more, unfamiliar information as a basis for expanded design decision-making, the discussion in this section deals equally with an established knowledge base and the attempt to add to that knowledge base by adding new information produced by research. The examples discussed in Chapters 4 and 5 both illustrate the problem of incorporating new information into a particular decision, and those examples will reappear in this section to illustrate the definitions being discussed.

Flexibility is achieved when, on a knowledge representation map, the knowledge recorded spans a large area on the map. In Figure 12, the region labeled "C" has low flexibility because it has only a limited extent in either level of detail or nature of information. The regions labeled "A" and "B" have high flexibility. As in the working definition given in Chapter 1, flexibility refers to flexibility in usage for design.

Two common modes of improving flexibility in knowledge are illustrated in Figure 12. One mode is to expand the knowledge in a way that increases its span by relying on connected information, as in the region labeled "A". Another is to rely on implicit connections between two separate areas, as in the region labeled "B".

Examples of these two flexures appeared in the previous two chapters. A case study contains narrative references to a wide variety of details and abstractions. A case study makes reference to a variety of academic disciplines. A case study contains both quantitative and qualitative information about the situation being discussed. Figure 13 illustrates how a case study can be used to achieve flexibility by connecting together information about failures, such as the mechanism of the failure and the data used to diagnose the failure by reference to the details of a particular failure. The case study illustrated in Figure 13 is the strain-age embrittlement case reviewed in the previous chapter; also diagrammed are some of the disciplinary areas which make contributions to the case.

The development of manufacturability data illustrates flexibility by relying on implicit connections between two regions in the knowledge

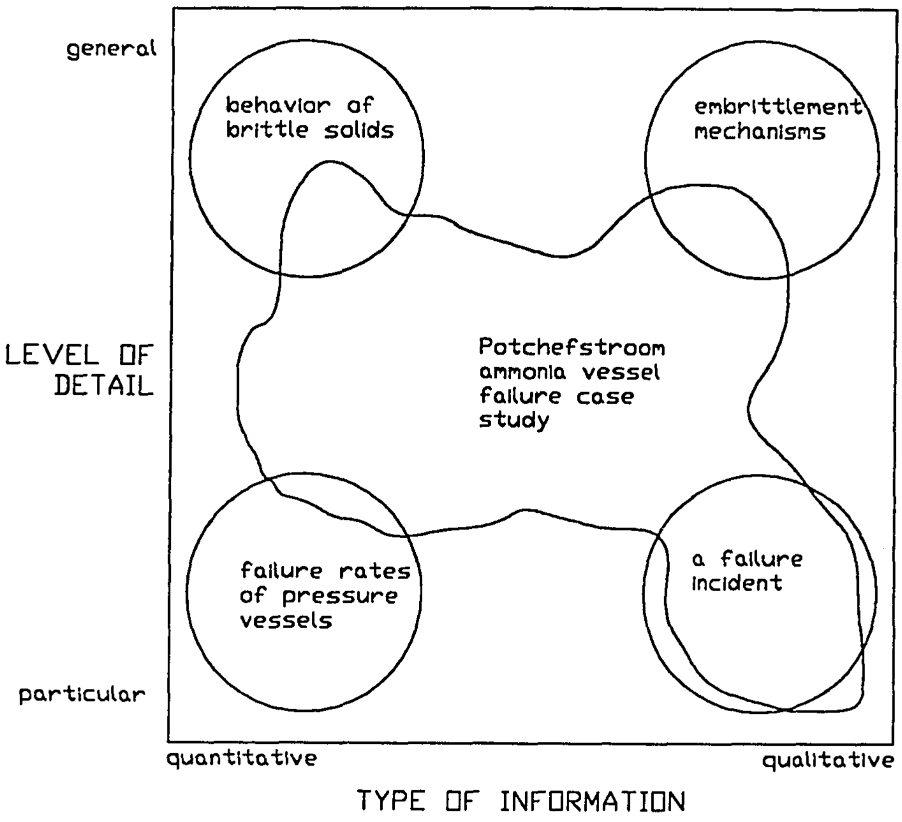


Figure 13

Knowledge Representation Map for A Case Study representation map. Figure 14 illustrates how machinability data is connected only implicitly to the more abstract machinability index discussed in the fourth chapter. For the machinability index, there are at least two links between the two areas, a quantitative link dealing with the empirical equation provided by Datsko (1966), and a qualitative link in the author's description of the physical processes involved in metal removal.

While flexibility refers to the extent or span of a region on the knowledge representation map, coherence refers to how strongly connected the region is. Thus region "C" of Figure 12, although of low flexibility, is of high coherence. Region "A" is also of high coherence, while region "B" is of low coherence.

In Figure 13, the high coherence of a case study is illustrated. In contrast, the machinability information diagramed in Figure 14 is of low coherence, revealing a location where further knowledge development can improve the general usefulness of knowledge about manufacturing.

Flexibility in knowledge is achieved by relying on knowledge with a broad span on a knowledge representation map. Coherence is achieved by assuring that the different aspects of the knowledge are connected together through explicit concepts which can be shown on a knowledge representation map; the highest coherence is a region which does not rely on links between separate areas.

A very common method of achieving coherence is suggested by region "C" in Figure 12: to remain in a narrow area on the knowledge representation map, sacrificing all flexibility. It is the author's opinion that the narrow, inflexible approach characterizes much engineering research today. Similarly,

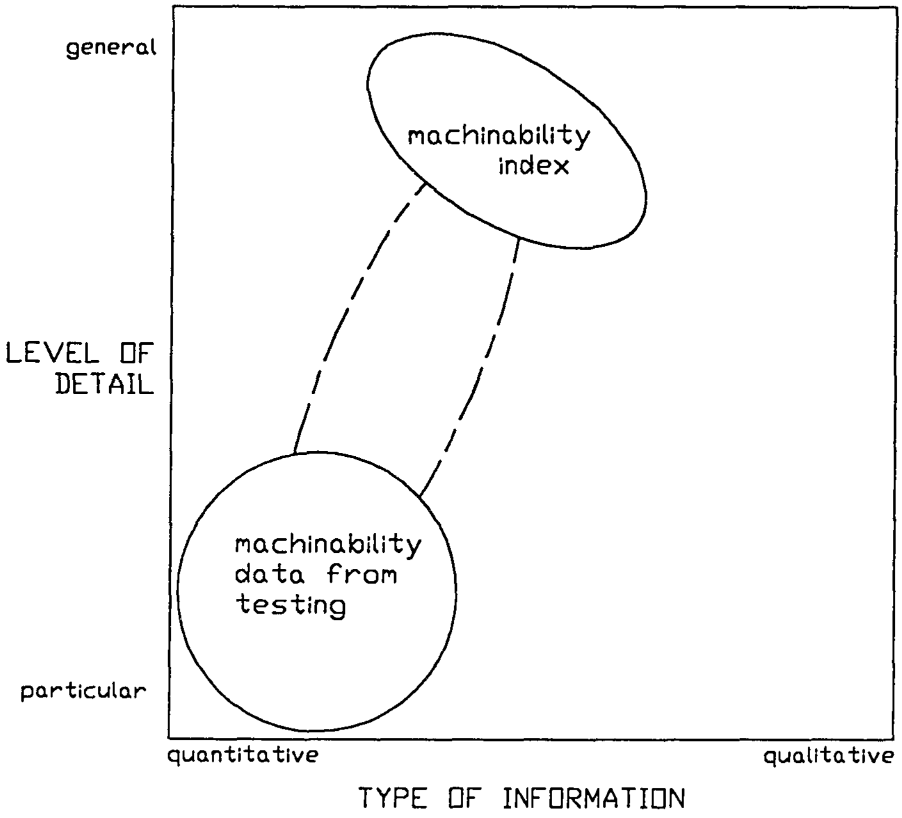


Figure 14

Knowledge Representation Map for Machinability one of the easiest ways to achieve flexibility is at the expense of coherence. It is necessary, then, to balance the two goals in order to achieve improvements in knowledge representation.

One can use the quadrant simplification of the knowledge representation map to simplify the definitions of flexibility and coherence. On a quadrant map, flexibility means simply the knowledge spans across more than one quadrant. Coherence means the crossing between quadrants is the result of strong connections within the knowledge being represented (region "A" in Figure 12), not on thready external connections (region "B").

On the basis of these refined definitions of flexibility and coherence, the remainder of the chapter contain discussions of some specific applications to improving knowledge representation for design.

6.3. Integration of the Design and Integration of the Design Process

Integrating computer-aided engineering systems is the subject of considerable research interest at present. At present, the goal of integration developers seems to be automating the process of moving a design from a CAD system to a computer-controlled manufacturing environment - that is, automating the process of communicating the design to the manufacturing facility.

Integration in engineering means communication. Peters (1987) considers the whole of computer aids to manufacturing as a problem of communication between human beings. Chorafas (1987) describes the primary function of CAE tools as enhancing communication between different people working on a project. Two key aspects are important: the emphasis on communications, and the fact that communications among human participants in the process (rather than the computers) is the paramount concern.

For the most part, integration in manufacturing is being pursued from the basis of large databases for improving access to information at a very detailed level (Yeomans, Choundry, and ten Hagen, 1985). By visualizing a knowledge representation map, one can see a large, detailed database has good coherence but has little flexibility. Flexibility, achieved by including more than numerical data in the information to be shared, improves integration by increasing the potential uses of the information for a given user, and as a result by increasing the number of potential users. Coherence must be preserved by developing conceptual links between the data and other forms of the same information. Coherence improves integration by supporting development of a context as a basis for reasoning about new problems.

Integration can be considered by use of a different view of the knowledge representation map, illustrated in Figure 15 and sharing a vertical axis with the map of Figure 12. The horizontal axis is orthogonal to the view shown in Figure 12. This map contains discrete listings of the disciplines involved in the knowledge base.

Integration can be considered in terms of which axis defines the principal direction of the interconnections within a knowledge base. Horizontal integration is dominated by connections between different subject areas. Vertical integration is dominated by interconnections between different levels of detail within a subject area. Integration of knowledge for the design activity is the same as developing improved coherence within the knowledge base.

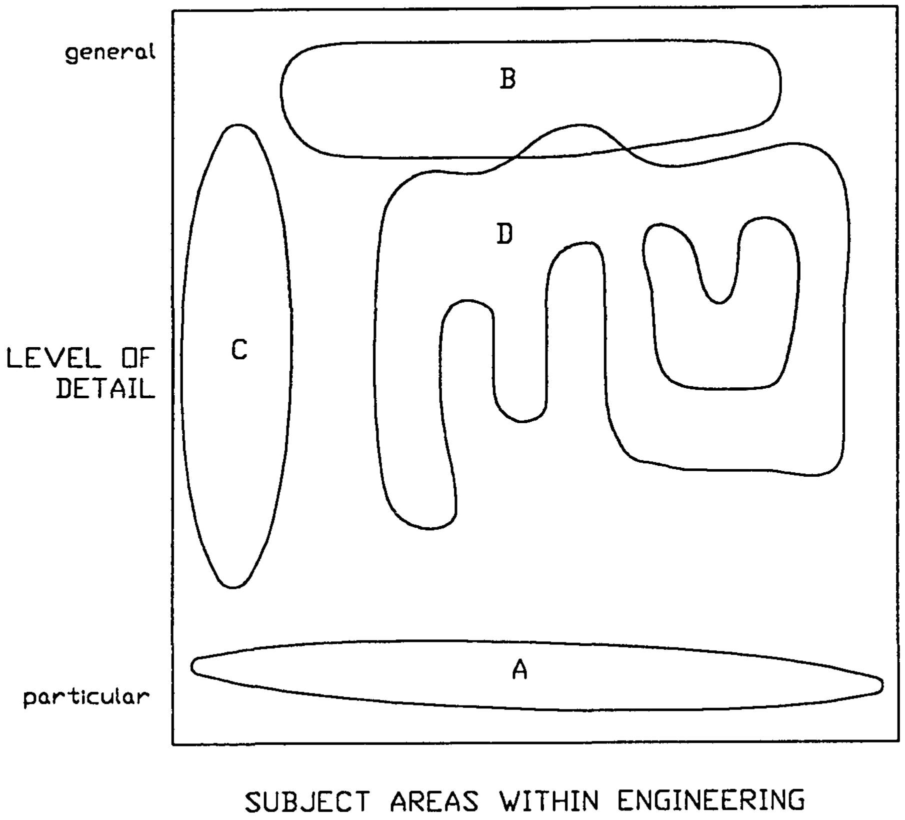


Figure 15

An Illustration of Integration

Horizontal integration, illustrated by region "A" in Figure 15, deals with connections between different subject areas which are important to the project. Region "B" is also an example of horizontal integration, although it is at a more general level of detail. The connection from material selection to the manufacturability indices developed in the fourth chapter could be represented by region "B", while the intended connection to material selection from the machinability data being used in databases for computer-integrated manufacturing is better represented by region "A".

Vertical integration deals primarily with connections between different levels of detail within a given subject area. Region "C" in Figure 15 illustrates vertical integration. The connection between a Material Failure Logic Model and the more detailed information about a failure mode is an example of vertical integration.

Both vertical and horizontal integration are important to achieve the communication improvements which will produce the gains in design productivity being sought. At present, development of integration approaches seem to focus on detailed, horizontal integration, illustrated in Figure 15 as region "A". In the author's opinion, improved vertical integration is the primary need in order to achieve economical, effective decision-making. Horizontal integration at the detailed level, illustrated by design for manufacturability as discussed in the fourth chapter and by region "A" in Figure 15, is expensive and painful to achieve and then loads down the designer with additional details. Horizontal integration at a more general level, as illustrated by region "D", is less difficult at the general level because the connections between disciplines is based on general principles of broad validity, not on the specifics of each area.

Contemporary work toward integration of design to manufacturing planning is focused on creating more access to the detailed manufacturing information (Yeomans, Choundry, and ten Hagen, 1985). This is an attempt to achieve horizontal integration, but a very restricted integration will result. It has been recognized that poor manufacturability in a design reflects poor connectivity between manufacturing knowledge and design. Research work in design for manufacturability is attempting to create more connections. Coherence can be improved through the creation of these connections, but at the expense of reduced flexibility because all of the new connections are being handled as details - in one quadrant of the knowledge representation map.

Coherence can also be improved through integration based on generalizations. Vertical integration can create a coherent and flexible knowledge base within each disciplinary area by making connections between detailed results of research and broader generalizations about the subject. Between subject areas, connections at a more general level improve flexibility and preserve coherence; doing so using knowledge that is economical to apply during design decision-making.

Horizontal integration, currently existing primarily at the detailed level, is appropriate at other levels as well. The design-for-manufacturability approach described in the fourth chapter supports horizontal integration at a less detailed, more abstract level of reasoning by connecting manufacturing concerns to material selection. Just as horizontal integration at more general levels is needed, vertical integration is needed through a variety of major emphases. The result is a network of horizontal and vertical connections within the knowledge base needed to carry out design. The network provides for increased flexibility and coherence in the engineering knowledge base.

The most important point to recognize is that horizontal integration between disciplinary areas is easier to accomplish, more economical, and generally more flexible at a generalized level. Therefore there is a need for two kinds of knowledge development for design: development of vertically-integrated knowledge within single disciplinary areas, and development of generalized, horizontal connections between areas to make the engineering knowledge base more coherent.

Improved connectivity within the knowledge base does not require a great deal of additional knowledge; for the most part the knowledge necessary to achieve better integration is already available. But available knowledge must be reorganized to suit the needs of the overall network, rather than just fitting into the particulars of a local issue. Specialized domains of knowledge tend to emphasize their own node in the network, often at the expense of the connections to the remainder of the network which makes that domain useful for design. In the worst cases, researchers operate as though there were no other nodes on the network at all, resulting in coherence locally but extremely limited flexibility. In Chapter 5, the large amount of available failure analysis information was contrasted with the considerably smaller supply of failure prevention information. This contrast is an example of knowledge which is extensive and coherent within its own localized area, but has not been connected well to other subject areas, and for that reason lacks the flexibility necessary for design. In the case of failure prevention, the result of poor connection to the general engineering knowledge base is persistent recurrence of preventable failures. In the case of manufacturing, the result is designs of poor manufacturability.

In addition to improving connections between knowledge domains, it is also necessary to restructure knowledge for application at different levels of abstraction, achieving vertical integration. The disciplinary knowledge needed is generally available, although an understanding of how it may be restructured requires further development. The restructuring must be guided by the needs of application in product development. The manufacturability knowledge developed in the fourth chapter is an example of such restructuring from knowledge which is already available.

Improvements in integration deal with making improvements in the flexibility and coherence of the engineering knowledge base. Integration is a convenient unifying concept, because it can be used as an omnibus which conveys much of the substance of current issues in design.

6.4. Communicating the Design

Integration of engineering, viewed as a whole, is the problem of improving the design as a device for communication. Ordinarily, it is assumed the design is the detailed part and assembly drawings, intended for communication between the detailing designer and the fabricator. This is but one, limited view of the range of communications required in the design. One should not imagine the design is recorded entirely in the drawing.

Another purpose of the design is to convey to users how the system is to function. Users do not generally require this guidance when the system operates correctly, so the design must incorporate information to support trouble-shooting and repairs. Often a technical manual is prepared to supplement the drawings; this manual contains primarily information generated by the designer.

A very important use of the design is to maintain corporate history. Some component types are used repeatedly, and for that reason the drawing of an existing, proven component can be adapted to new uses with considerably less expense than developing an all-new component. This is commonly referred to as the need to avoid reinventing the wheel. In the past, adapting existing designs to new uses relied on an individual engineer's experience with a component type, rather than on recording the experience in a file belonging to the organization.

As noted in the fifth chapter, major selling point for CAD systems is that they help the organization avoid reinventing the wheel (Gaertner, Nieva, and Argoff, 1986). It is true that a drawing on the CAD system is readily available for adaptation to new purposes. The U.S. Navy (Webber, 1986) is now providing magnetic media copies of CAD drawings to its offerors as part of the technical information in bid packages for design and construction of its ships because the drawings can be adapted by a bidder for cost estimating, for evaluation, or for proposing modifications.

Cognition (1987), and similar tools recently introduced to the market, deal in part with avoiding prior problems. The design the Mechanical Advantage workstation records is not just the drawing, it also includes the formulas and calculations used to obtain that drawing. The Mechanical Advantage system provides partial vertical integration, because a sketch developed on that system can be transferred, through an industry-standard data structure, to a CAD system more capable of recording the details as they are filled in. The CAD system might in turn be used to communicate the geometry to a CAM system.

An effective, integrated design tool must provide the means to record not only the finished part and assembly drawings, but also the designer's intent and concerns as part of the design file. Reinventing "square wheels" can be reduced if notations about the problems which were considered and avoided can be recorded in a reasonably compact form. Compactness is important for two reasons: there is a limit on how much more information can be stored in the design file, and there is a limit on how much more information can be handled by the engineer who wishes to adapt the design to a new use.

Another aspect of horizontal integration is distribution of information to the many actors involved in the different aspects of product development (Yeomans, Choundry, and ten Hagen, 1985). Data and knowledge important to the design exist in a variety of forms and in a variety of locations. Vertical integration of the knowledge base provides the avenue for economic communications between the members of a design team. In developing vertical integration it is necessary to consider how the knowledge at one level may be translated to another level. An example of this exercise is the manufacturability approaches developed in the fourth chapter, where the goal was to take detailed information and translate its vital elements into a broader context.

It is also necessary to coordinate applicable data about the design and about the manufacturing situation for use at other levels. One approach to this can be based on the case studies applications discussed in the fifth chapter. The purpose of the case study systems was to take experience with in-service and manufacturing failures and make them useful during earlier decision-making. The original intent is to provide guidance for design checking, but an important secondary goal is to give enough feedback (from prior experience, other persons on other projects) to engineers that they can learn to avoid such problems during design rather than correct them after commission.

6.5. Context for Design Decisions

The context created by a case study is a primary reason for the success of the case study basis for design review, which was discussed in Chapter 5. For the purposes of engineering design, context can be defined simply as the information which allows a generalization to be correctly interpreted and applied to a specific problem, which indicates how coherence and flexibility are both necessary to support a context. In other words, the engineer's view of the problem must be broad enough to understand the how new information relates to the problem. In a sense, context refers to the connection between what the engineer needs to know about the specific design problem and what he already knows about other, related problems and about methods of problem-solving. In this section, two applications of context will be discussed: the design checking ideas which were examined in the fifth chapter and the problem of introducing new technologies to a designer.

The connection between a problem to be solved and the knowledge base available to solve it is influenced by flexibility and coherence in the knowledge to be used. Limits on these two properties limit an engineer's ability to apply the knowledge because the connections needed to obtain a context may be incomplete, missing, or too difficult to create.

Many modes of failure are not familiar to designers, and to locate and eliminate a potential failure which is not familiar, the engineer requires a connection between the unfamiliar, new information and the knowledge already available. The case study review, discussed in Chapter 5, provides an example of a context-setting method.

An effective way to create a context is to provide an example. In the case study reviews, the example is the narrative of a particular failure case. The case study description contains not only the example, it also makes reference to the mechanism of the failure, giving the root causes and examples of them, and to the usual progress of the failure. This provides the abstract information about an unfamiliar failure mode in the context of a concrete, recognizable example, providing a path by which unfamiliar information can become usable.

Many new technologies and new ideas are developed with the intent that they will improve products. Efficiency in a power plant is improved primarily by increasing the difference between the coldest and the hottest locations in the plant, making development of high-temperature materials desirable. Improvements in both aerospace and ground vehicle applications rely heavily on increasing the load-carrying capacity of a vehicle without increasing the weight, leading to development of high-strength and high-stiffness materials such as fiber composites. However, these new materials have properties different from others which are more familiar to engineers, making it difficult to use them effectively.

Fiber composites provide a useful example of the problem of introducing new information. The author's reading of a sampling of the literature on composites, such as Agarwal and Broutman (1980), indicates that researchers seem to concentrate on why composites are different from metals in their mechanical behavior. Differences might be important when considering the details of an application, but when trying to decide if a composite will serve where the engineer might have put metal, the ways in which a composite behaves like the metal it might replace are important. The similarities between a composite and a metal are the context, which allow the engineer to reason about candidate materials on a basis of equivalence. Even if the differences are important, similarities create the connection to what the engineer already knows, allowing the exceptions to be treated properly.

Without a context, the engineer is constrained to making a decision based on incomplete evidence. This has in the past led to problems. The bankruptcy and reorganization of Rolls-Royce in the early 1970's can be attributed largely to a decision to use composite blades in the compressor section of a jet engine being built for large airliners (Eddy, Potter, and Page, 1976). The new directors of the reorganized company made an equally underinformed decision to eliminate composites, and replaced in metal some plastic and composite parts which were unrelated to the problem blades.

In the author's experience with tensile tests on glass-polyester composite specimens, there are many important similarities between composites and the more common isotropic materials. For example, a random-fiber composite exhibits linear, elastic behavior up to a limiting stress, when a decrease in the slope of the stress-strain line is observed. Beyond that point, there is a non-recoverable component of damage as well as a recoverable elastic component. Even though the mechanisms are entirely different, the author finds a useful analogy to the tensile behavior of a work-hardening metal.

A directional composite, loaded in the direction of the fibers, will exhibit linear elastic behavior right up to failure of the specimen. It seems reasonable to classify such a material as high-strength, but brittle.

Even the compromise between strength and ductility in metals has an analogy. A directed-fiber specimen has higher strength, but is brittle when it fails. A random-fiber specimen, identical in fiber and matrix constituents and volume fraction of fibers, has considerably less strength but exhibits the ability to accumulate damage prior to failure, which one might call ductility. The same relationship is true in metals, for instance when comparing annealed steel with quenched steel. Annealed steel is of lower strength, but greater ductility. The steel can be quenched to increase the strength, at the cost of reduced ductility.

A current project of the author's is the design of a remote-controlled vehicle. The chassis design is critical because of weight limits on the vehicle, therefore fiberglass construction is being considered. It has been very important to account for the fact that the strongest glass-fiber composite is also the most brittle. The trade-off between high strength and robust, "ductile" performance has been the subject of great deliberation.

This example illustrates the uses of context. An engineer who has no familiarity with composite materials can begin thinking about a possible application based on reasoning by analogy to metal properties. The engineers working on the remote vehicle have no extensive experience with composites, yet they are making decisions which advance the design. Later in the process, with the context clarified by the particulars being developed in the design, the engineer is in a position to examine more detailed information about how that composite is best applied. For the remote vehicle, it is planned that the structural detailing will be carried out in consultation with a subcontractor who will be laying up the fiberglass chassis.

Any new technology, to be usefully applied, must be presented in a context which gives connections to the existing knowledge base. This is necessary to allow the engineer to reason about how the new technology can take the place of an older idea, as discussed above. Connections to existing ideas are also necessary because the new technology must be used in connection with a large number of other technologies when it is applied in a design.

In general, one may assume the design engineer reading a report of recent research is prepared to understand the information being presented. The engineer has the prerequisite foundation of knowledge. The difficulty lies in the assumptions that are made about how that knowledge will be used. Engineers are not necessarily able to begin using new information as soon as it becomes available, and this is not a weakness on the part of designers. The main improvement needed in knowledge representation for engineering design is to consider, explicitly, how that knowledge relates to the decisions which must be made. In other words, work at improving coherence and flexibility is needed.

In recent years, engineering literature has begun to reflect a growing dissatisfaction with the way knowledge is provided to engineers. Nadler (1981) criticizes researchers for making the assumptions that producing new information is an end in itself, and that engineers will be able to apply new knowledge in design as soon as it becomes available, regardless of its form. Barrett (1981) makes the point that research results are usually reported in a form not immediately usable by designers.

Seireg (1987) expresses the opinion that America's engineers are good at developing new technologies but not at following through to applications. This difficulty , Seireg says, is often expressed by researchers as one of engineers in design positions failing to capitalize on new technologies. The author's opinion, largely supported by discussions with engineers in industry as well as much of the commentary in the literature, is that the main problem lies with research results being presented in ways that limit the design uses of the knowledge. Engineers are not necessarily able to begin using new information as soon as it becomes available, and this is not a weakness on the part of designers. Engineers are unable to capitalize on new technologies because they are provided only with a few incomplete glimpses into a new idea, not because they are incapable of applying an idea once it is clarified. The distinction between failure analysis and the knowledge needed for failure prevention in design, discussed in the fifth chapter, provides an example of a research approach in which the reports of results do not match designers' needs. The mismatch is because the failure phenomena are studied in ways that emphasize how they behave once they occur and how to create them so they can be studied, while designers need to recognize potential for failures and prevent their occurrence.

Experienced engineers can still have problems with the knowledge they apply. A manifestation of these problems are recurrence of recognizable, preventable errors in the design. Continuing development of the engineering knowledge base is necessary to improve the effective use of knowledge even after a technology has become familiar.

It can at times be difficult to distinguish between data, information, and knowledge. The discussion in this chapter helps to clarify the distinction. Data always falls in the lower left quadrant of the knowledge representation map of Figure 12 - always detailed, always quantitative. Information tends to fall in only one quadrant, therefore data is a form of information. In contemporary usage, information typically means detailed, but non-quantitative material. Knowledge, in the author's opinion, is distinct from information because it falls across more than one quadrant. Therefore the properties which characterize knowledge of general value (as opposed to information which has value only in particular, constrained settings) are flexibility and coherence. Unfortunately, by this definition, most of the results of knowledge-producing research is stated in the form of information, not as useful knowledge.

6.6. Conclusions and Recommendations

At this point it is possible to summarize four conclusions from the research. These deal primarily with the question of knowledge for design use, but also have an important influence on how engineers develop and use design tools, which will be examined in more detail in the next chapter.

1. Researchers, who are knowledge producers, must consider when they record it how their product will be used. Just as the developer of a machine must consider its users, developers of knowledge must consider users' needs as well.

   Perhaps the best phrase for describing the improvement needed in research reporting is fitness for purpose. Knowledge must be stated in forms which make it useful for design purposes. The dual goals of flexibility and coherence provide guidance on improving the design value of knowledge.

2. In addition to knowledge-producing research, there is a need for engineering research into knowledge representations for design. This type of research can enhance the fitness-for-purpose goal. In the near-term, such research will reduce already-existing information to manageable levels, and reformat it into representations which are better suited to design purposes. In the longer term, engineering knowledge research will clarify the relationship between knowledge, skill in applying knowledge, and design decision-making, and in particular will instruct knowledge-production researchers on improving fitness-for-purpose in their products. Another important, and frequently neglected, area for research is the set of connections between the individual disciplines, and how these interactions affect application of the knowledge areas during design.

3. Knowledge for engineers, and knowledge in general, cannot be stated only in abstract terms, nor strictly as details. A context, defining the nature of the problem addressed, aids comprehension and makes it possible to reason about how well the knowledge in the research report will serve a design situation. Context is also necessary to assure knowledge about a new technology can be connected to the existing core of engineering knowledge, which is necessary to be able to use the new technology effectively. Context for a generalization is achieved by preserving coherence as flexibility is improved.

4. Integration of the design process can be described simply as improving the communication connections between different issues

within the design process. Although current efforts at integration are worthwhile, they are limited to horizontal integration between a limited number of disciplines, and at a detailed level. Horizontal integration at more general levels of detail are also important and will probably be more effective. Vertical integration is an important means of improving both flexibility and coherence in knowledge for design.

This chapter has drawn some general conclusions about improving the management of knowledge provided to the design process. In addition to further knowledge-generation research, there is a strong need for research carrying out knowledge structuring. The next chapter will apply these ideas to the particular problem of building tools to assist in design decision-making.

Chapter 7
Computer Aids to Engineering

In the previous chapter, the author developed some general ideas about the forms of knowledge required in support of design decisions. In this chapter those ideas are applied to the problem of building tools to carry out decision support.

First, a discussion of the design tools currently available is necessary. This provides a baseline of what decision supports are already available. These tools are in general limited to computer-based systems, and include the graphics CAD systems, advanced numerical analysis tools, such as the finite element method, integrated databases, and expert systems.

An examination of identifiable niches where further development of tools would be appropriate will follow the review of design tools. The discussion will focus on those problems for which support tools appear to be feasible. Some recent introductions to the CAE market which reach toward meeting those needs will be considered.

While localized design and subsequent detailed evaluation are reasonably well-supported, the discussion of this chapter will show that global design is the area receiving least support. In the second chapter it was shown major commitments are made during early, global design which will influence the success or failure of the project. There are, however, approaches which can be used to improve support for decisions which must be made early in the project.

7.1. State of the Art in Computer-Aided Engineering

Nearly every effort to improve engineering productivity grows into an acronym beginning with the letter "C", for "computer". In this section the author will examine the established tools, all of which incorporate use of the computer. This includes graphics and drafting systems, databases, detailed numerical analysis methods, and expert systems.

The beginning of computer aids to engineering was the graphics system, now called a computer-aided design (CAD) system. Although originally conceived as computerized drafting tables, expansions have added capabilities in several directions. The primary value of these systems is the ability to modify, and manipulate drawings with minimal effort by the draftsman. These abilities lead to many of the features for which modern CAD systems are appreciated: assembly and interaction drawings, simplicity of corrections and modifications, recall of old drawings for adaptation, and communication with automated manufacturing systems.

CAD represents a substantial contribution to the portion of the design process which is concerned with detailing the geometry of the components and communicating that geometry for a number of different purposes. But as Foar (1984) observed, "the new-found power in the drawing office has enabled drawing practices to be automated but the job of fundamental engineering design is still the same as it was." A similar comment was made by Knight (1983). The support provided to this single aspect of design should not be confused with support to all of design.

There is a growing emphasis on the use of video images which offer a three-dimensional view of the design. Witwer (1989), for instance, claims substantial reductions in design time by the use of Silicon Graphics workstations which permitted design in three dimensions. All of the examples cited in his article, however, dealt with very detailed decision-making, taking place late in the design process: pipe routing in a power plant, assembly and fit of automotive doors, and aircraft cockpit control panel layouts. It is important to realize the workstations where these three-dimensional images were manipulated for design are still expensive enough that a limited number of them will be used for engineering.

In early decisions, exact geometry is not important. What is important is quick and simple depictions of the geometry. In fact one technique for enhancing creativity among engineers is called "rapid visualization" (McKim, 1967), which relies on quick sketches of the ideas being developed. Tomiyama and Yoshikawa (1985) observe that the ability to handle quick, rough sketches is quite limited in most CAD software. Christie (1985) explains that detailed pictures may seem attractive and may convey more feeling or atmosphere, but schematic or cartoonized images have been shown to convey more useful and more memorable information in many cases. However, early design does not necessarily mean inexact geometry; the point where the drawing becomes more exact is determined by the problem context.

The drawing is only one of several important aspects of the design. Particularly in the early stages, the notes and comments in several different pencil colors on the face of the drawing are as important as the sketch itself. The combination of words and picture together convey far more than the picture alone. The cumulative commentary obtained on a drawing as the original idea is reviewed - a question is noted, its answer provided, which leads to another question - is a valuable tool during design. These comments provide vital information, and are without question part of the design which needs to be recorded. For this reason, a Vice President for advanced projects at a large aerospace firm (Kuzela, 1989), complains that the proliferation of CAD systems has reduced creativity among the engineers he has dealt with.

Another issue in early design is also not supported by conventional CAD systems. This is the question of whether the idea is feasible, which Asimow (1962) calls physical realizability. This evaluation does not deal with geometry alone. The discussions in Chapters 4 and 5 covered issues of great importance to physical realizability while ignoring geometry.

Another acronym enjoying popularity at present is "CIM", which stands for Computer Integrated Manufacturing. A major thrust of CIM is making sure everything that needs to be known is available through similar means, in particular through a large database containing information about the product being developed and the manufacturing processes available for use to make components of the product. The second type of tool to be discussed involves making available to the designer the material properties and tool capacity and allocation databases used by manufacturing planners. This is referred to as an integrated database.

The machinability information available from a typical database deals with specific alloys and heat treatments connected to particular metalcutting processes (*Machining Data Handbook*, 1972). Such a database for a large manufacturer might fill several printed volumes. While anticipation of machinability is important, the detailed database approach involves a greater degree of effort for the engineer. The approach discussed in the fourth chapter, in contrast, provides a more economical approach especially for early decisions, based on minimal knowledge and minimal additional effort to make a suitable comparison.

The integrated database approach to design for manufacturability concentrates on horizontal integration at a detailed level. When finalizing the details of a product design it is indeed appropriate to make decisions based on detailed information about how it will be fabricated. A late, localized decision can be made on a detailed level because most options have been ruled out by prior decisions, limiting the amount of searching required to obtain the necessary information in the database. But providing the engineer with a large supply of additional information during the entire design process leads to losses, not gains, in engineering productivity. In the discussion of integration in the previous chapter, it was shown that detailed, horizontal integration is less economical for global decisions than an approach emphasizing vertical integration with the horizontal connections achieved at a more general level.

A database can be a useful tool for design, provided the information the engineer seeks can be accessed quickly, in a form appropriate to the needs of the design decision. A flood of detailed information requires extra time to sift through, adding to the engineer's burden. A database appropriate to design would have flexibility, incorporating several levels of detail. The database must include several successive levels of abstraction above, the specific, detailed data; each level providing a broader, less specific view of the information stored there. Depending on where in the design process the decision is being made, at least one of the levels should mesh with the needs of the engineer.

The third class of tools to be considered are detailed numerical simulation methods, which are often called analysis methods. Of these, perhaps the most prominent is the finite element method, but there are other examples as well.

An important recent criticism of numerical tools for analysis was given by Smith (1986). He has worked for several decades in turbomachinery design, and has participated in the entire history of computer-based design tool development in that area. As computers became more capable and tools grew in size, scope, and complexity, interpretation of the results grew increasingly difficult. This leads to the problem of "garbage in, gospel out" (Roszak, 1986, credits this expression to Ashley Montague). Checking the results is increasing in importance, but at the same time increasingly difficult.

An analyst for a heavy equipment manufacturer explained that he will not begin modeling until he has carried out simple hand calculations. In his view, the finite element method has value in comparing alternatives, but the model must be thoroughly tuned based on experimental evidence before it can be used for absolute predictions. For that reason, the advance calculation by hand, however crude, is necessary to allow him to determine whether the model results are valid.

Finite element models can be used for absolute prediction, provided the validity of the model is carefully defined. After an extensive physical test program which required years, an engine manufacturer relies on the finite element method for absolute predictions in engine flywheels, and no other parts. This component was the first target for definitive numerical simulation because flywheels are physically tested to destruction, and the shrapnel from a flywheel failing makes for an expensive test facility. Other common components, such as engine connecting rods and steering mechanism links, are planned for similar testing and model correlation, which will require many more years of work to assure the numerical model is an accurate depiction of physical reality.

Detailed analyses provide only limited feedback to the designer. In the beginning of Chapter 3, the author made a distinction between direct problems and inverse problems, observing that carrying out design is not accomplished simply by inverting an analysis method. The limits on stress analysis methods provides an example. The model can show predicted values (of stress, for instance) to the designer, but if the evaluation indicates the design requires a change, neither the modeling method nor the output from that method is of much use in determining what changes are appropriate (Cross, 1942).

Numerical simulation methods, such as the finite element method, supports a final evaluation of a final geometry, but predictive feedback on improving the geometry is not readily available. The original geometry must be completely defined to provide the necessary input. The simulation methods require a fully defined geometry in order to operate and are therefore used primarily to compare a limited number of detailed alternatives. Like a detailed database, most numerical modeling methods generate a flood of detailed data, some of which may be useful. A burden is placed upon the engineer to seek a valid interpretation of this volume of numbers.

Numerical simulations have their uses and will remain important to engineers. But like the other tools being discussed they have limited value as design tools in their present forms, primarily due to the emphasis on details. Numerical simulation methods do not ordinarily provide information of predictive value, they are only capable of generating feedback on complete designs.

The fourth type of tool which requires consideration is the expert system. On their introduction into engineering several years ago, expert systems experienced a rapid expansion of popularity. Now, interest in expert systems is more cautious. While there are some successful demonstrations of expert systems and there appear to be some areas of useful application, it is not obvious that expert systems will make a large contribution to engineering design.

Some of the concerns about expert systems deal with the question of initiative in the system. Although an individual engineer may be willing to have certain decisions made for him, we are not in general willing to accept decisions which are made without an engineer's oversight. Gero (1974) made the observation that society will always want an individual available to take the blame. Yet expert systems take the initiative from the user, with the data required and the questions to be answered driven by a strategy fixed in the system's software. These question-and-answer sessions may seem to the user laborious, rigid, irrelevant, unnatural, and inefficient (Kidd, 1984). A system which leaves the initiative with the user would allow him to volunteer the information which seems pertinent easily and rapidly, with the system carrying out most of its reasoning based on this information.

There are many applications of expert systems in engineering. These may be characterized by problems which are well-understood and narrow in scope, and solved by well-established solution methods. For instance, an expert system in use at Du Pont for quality control of textile fiber production

(Horn, 1989) interprets the trends on process charts and advises on the necessary actions to keep the process within parameters.

But Horn (1989) also notes, "an expert system is not a substitute for a human expert. Many expert system projects have failed precisely because the complex problems they were intended to solve were not fully understood. That understanding can only come from humans." Of all engineering activities, design is the least well understood because it requires intangible aspects of cognition, such as creativity and intuition, as well as the more definable aspects such as analysis and explicit rules. Even if the methods used to solve design problems were known, the design problem itself is rarely narrow in scope, and never well-defined. The author is not convinced design methods are knowable at all under the current state of thinking about design theory and expert systems. Expert systems to solve design problems are not suited by all three of the criteria cited above: design problems are not well understood, have wide and largely unpredictable scopes, and rely on poorly understood solution methods.

Expert systems are suited to diagnosis tasks, where the data required by the expert system is generally obtainable. There are many applications for which a diagnosis seems worth the effort of feeding information to the system. But expert systems are inherently single-issue specialists. Just as human experts in a single area have only limited value in a design situation, expert systems are also limited because in design the breadth of connections between issues is as important as the individual issues themselves and more important than any one isolated issue. No plausible solution has been found to the limit which prevents construction of larger knowledge bases, to accommodate more than one realm of expertise at once (Buchanan, 1982). Even if there were no limit to the size of an expert system, it has yet to be shown that the connections between separate issues can be handled in the same way as the knowledge for an issue alone (Winograd and Flores, 1986). Woods (1986) cautions that performance of a decision-making system in coordination with other systems cannot be predicted from its operation in isolation, indicating an important reason expert systems for design must be treated with caution.

There is a more fundamental objection to the long-term value of expert systems in engineering problem-solving. The context set by the problem to be solved is the primary consideration in choosing a solution method in design. Yet construction of an expert system is predicated on the assumption that a structured collection of rules, which contain context-independent knowledge, may be used to solve problems (Hayes-Roth, Waterman, and Lenat, 1983). Therefore the basic assumption of context-independence behind expert systems seems ill-suited to the context-dependent knowledge and problem-solving approaches of design. In the previous chapter, the author described the importance of flexibility in the structure placed on the knowledge, since the structure used constrains the ways in which the knowledge can be applied; Expert systems have little flexibility. Stromboni (1986) and Hillman (1985) express similar concerns about expert systems.

Koen (1985) believes that a majority of engineering decisions are based on methods that have no global validity, and are therefore used only in appropriate problem contexts. In the sixth chapter the importance of context in learning and applying new knowledge was discussed. Winograd and Flores (1986) agree that context is important, as do Olsen and Anderson (1986). Dreyfus and Dreyfus (1986) object to the expert system postulate of context-independence, arguing that the context in which the knowledge was originally learned by an expert, and the context of the problem to be solved, are as important as the abstract version of the knowledge which is the ordinary focus of expert systems research.

The most successful design applications of expert systems, such as the packaging design system of Brown and Breau (1986), deals with standardizable, well-constrained problems only. More importantly, the sub-problems assigned to the automated design system must be clearly isolated from the remainder of the product. Brown and Breau's shipping container for computer components could be designed in complete isolation from and after the design of a component. Routine, isolatable problems are the only area of design in which it is clear that expert systems can succeed at automating the design process. This is not to say that routine, standardizable design problems are necessarily easy to solve. As discussed in the first chapter, the nonroutine, ill-defined aspects of design are the ones where greatest support is needed. Expert systems may have their place, but they, like the other tools discussed in this section, are limited to fairly routine, well-defined areas of design and limited to late, details-oriented decisions.

One can consider global and localized phases of design, which were discussed in the second chapter, as concerned with pre-details and details, respectively. In the second chapter the author discussed the distinction between performance and quality concerns in the design. The design tools discussed above concentrate predominantly on performance and details, with considerably less support to the performance, pre-details combination or to any quality issues. This is illustrated in Figure 16, which is the author's assessment of the state of computer-aided engineering.

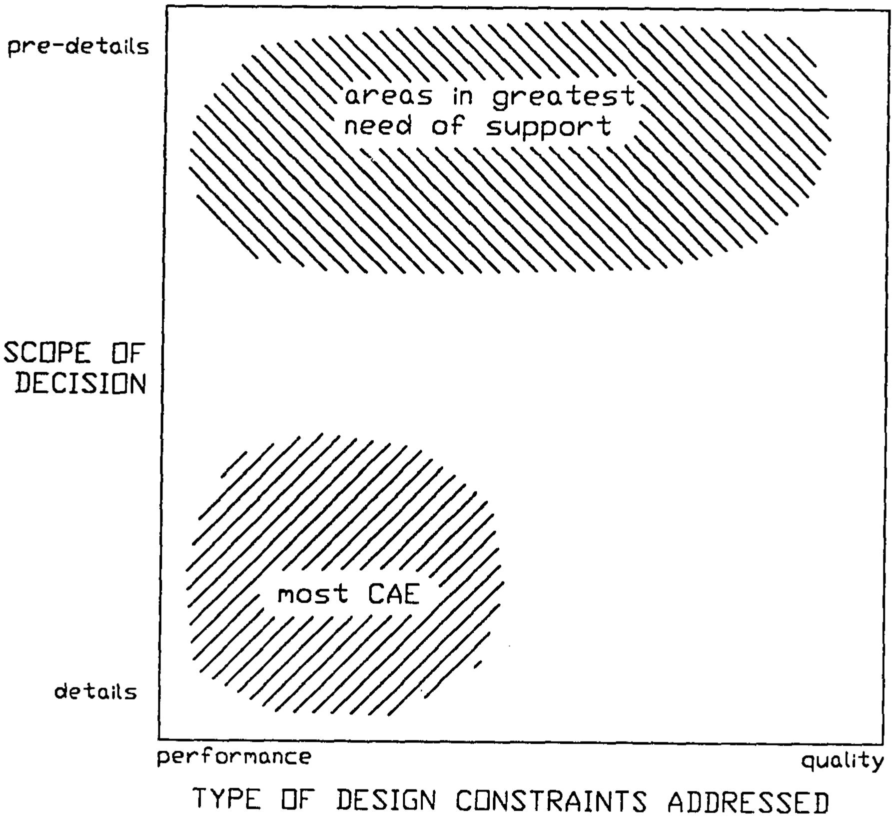


Figure 16

Support Available from Current CAE Tools

7.2. Design Automation

Design automation is currently a subject of considerable interest. In general, the term is taken to mean that some portion of the design can be produced by a computer, without human intervention (Winston, 1984). Originally, the intent was production of programs which were as intelligent as a designer for selected problems (Brown and Breau, 1986). As the predicted time to achieve such a tool continued to increase, that ideal was set aside, to be followed by the argument that considerably less intelligent approaches could succeed because of the computational power available in modern computers (Hatvany, 1985).

Like the "islands of automation" in the factory, which reflect the locations where manufacturing automation was easiest to implement, there are island of orderliness in design where design automation may succeed, locally, but in relative isolation from the remainder of the product development process.

This section contains an examination of the ways tools can assist an engineering designer, and a look at how the tools discussed in the preceding section fit into these uses. It has been observed in prior chapters that tools available today emphasize a limited subset of the types of problems which engineers face as they undertake design. Three major problems with the idea of design automation will be discussed in this section. First, not much of design overlaps with the computer technologies available to automate decision-making. Much of the advances made in design tool development are engineers adapting technologies which may serve a need; very little of design tool development is based on a direct understanding of what designers need. Second, there is a distinction between a decision and the result of an evaluation. It is not clear whether computers are capable of making true decisions in the ways humans make decisions. The third problem relates to the second, dealing with the question of how the computer's work can be checked (and whether it is checkable at all). These issues are reflected in the earlier discussion on the design tools currently in use.

The first objection to design automation to be discussed is the poor overlap between the needs of engineering designers and the capabilities of contemporary computer-based tools. Weiner (1985) comments the ready availability of new computers and new software concepts has brought about a troubling new concept, which he describes as "let's just add one more computer." Although Weiner's comment deals with automation in aircraft cockpits, it applies to engineering as well. A design group manager in the heavy equipment industry comments that many managers - even those with prior technical experience - are beginning to expect "push-button engineering", with rapid and error-free designs coming from the computer, rather than from the engineer. This was anticipated by Holland (1978), who voiced caution about the long-term value of many of the innovations introduced through the availability of the computer. Although several flashy new tricks are possible, these tricks do not necessarily constitute an improvement of substance to design, and in some cases they distract from design.

There is a distinction between software and knowledge, and between all the knowledge needed for design and the knowledge which can be coded in terms suitable for computer use. As an example, consider the problem of interface design for a software tool. One might assume that the on-screen display and how it is manipulated represents a majority of the interface design problem. But the interface is a communication problem, and communication lies not so much in the particulars of the interface presented to the user as it does on a shared understanding of the ideas to be conveyed - a context. In the sixth chapter, it was shown that integration and communication rely primarily on a coherent knowledge base which allows new concepts to be conveyed in the context of established ideas. Successful use of a computer program for fracture evaluation relies more on a grasp on the principles of fracture mechanics than on the ability to manipulate on-screen menus. Begg (1984) observes that automation of physical processes was difficult until the physical skills necessary to carry out those steps were analyzed and clearly statable. A similar clarification of the mental skills required for design is necessary if design is to be automated, but is a problem orders of magnitude more difficult than evaluation of physical skills (Price, 1985).

Foar (1984) observed there are many problems which cannot be solved by computational power alone. Design is one of those problems. If ordinary application of a routine procedure can resolve the situation, it hardly represents a problem (Kantowski, 1981), yet computer-based methods are best suited to routine problems (Brown and Breau, 1986). Christie (1985) believes that engineering problems are necessarily the nonroutine ones. Design is an example of a class of problems solved more effectively by finesse than by brute force. There are few design problems which are amenable to automated solution through the ways computers are now being used. In part this problem is driven by the fact that engineers are still working with what tools software developers provide, instead of specifying the needs that new software must meet.

Knight (1983), in a survey of design tools, concluded that very few of the design systems available actually contributed to creation of the design.

This is not necessarily an undesirable situation. The concluding section of this chapter will outline approaches to design tools which rely on them as tools - extensions of particular human abilities, controlled and coordinated by the human engineer.

The second objection to design automation, in the terms of Winograd and Flores (1986), deals with the nature of commitment in human communications and decisions. They argue that a human conversation cannot be replicated by a similar dialogue between computers, even if the wording of the transcripts is identical, because at the root of the human conversation there are commitments being made - that such an idea will solve the problem, that this person will carry out certain actions. The real communication in a conversation is commitments to action. However "intelligent" a computer is programmed to behave, it cannot make commitments to action because it probably lacks the means to carry out that action, and because a commitment is made as much on an ethical as a logical basis.

The same argument can be applied to a design evaluation. Price (1985) believes "it is important to recognize that design decisions are inventive actions rather than simply products of analysis." The output from a finite element analysis is not a conclusion, it is evidence used to obtain a conclusion. A stress calculation is not an end goal for the engineer, it is an intermediate step required to reach the goal of evaluating the adequacy of a load-bearing component. The same is true for any other analysis. While a computer can produce numerical results, can it interpret them? Where the failure mode can be evaluated numerically or logically, the computer may even be capable of making the comparison against a given criterion. The computer is not, however, capable of reasoning about which failure modes require evaluation. Computers can generate evidence, computers can manipulate evidence to alter its form, but computers cannot use a mix of kinds of evidence as the basis for a decision.

For example, assume a computer has been programmed with the American Society of Mechanical Engineers Pressure Vessel and Piping code standards for welding, as well as the American Welding Society structural welding code, and the welding recommendations of the American Institute of Steel Construction. (Even if such a computer is hypothetically possible using current technology, it is not feasible because of the practical working limits on expert systems.) One might assume that, with all that welding information available, the computer could easily design and evaluate welded structural joints. But even within a single code, there are variations and exceptions; there are variations and contradictions in some cases between different codes. Which code should apply depends heavily on the problem context - how and where the weldment is to be used. After all the welding information has been programmed, considerably more work will be required to program the reasoning required to decide which evaluation standard to apply, and how to apply it.

One may argue that the limits on computer capabilities are solvable problems, but their solution depends heavily on understanding how design decisions are made. As mentioned earlier, this understanding will require more work before it is achieved. The discussion in the section to follow will emphasize the use of design tools as evidence-management systems to support design decisions, not as decision-making devices.

The third objection to design automation is the additional engineering workload generated by automation. If we elect to allow a computer to generate a design, or a portion of the design, can we trust the result? Campbell

(1984) thinks we cannot, even after the software has been tested and tuned. Gero (1974) expressed a similar concern. This is related to Winograd and Flores (1986) argument about commitment: when human engineers says the design will work, they are making a commitment that they have made every necessary evaluation and judgement. An engineer's commitment incorporates some degree of personal and professional risk, and engineers must be prepared to face the consequences if they are wrong. When a computer says the design will work, it is indicating only that the design meets those criteria which it has evaluated, and it cannot make a commitment. The distinction is subtle but important. Although a computer is capable of mimicking some forms of human reasoning, it is not capable of reasoning like a human engineer would. We tend to doubt the opinions of an engineer with little design experience; we will not trust the results of computer-generated design decisions even more because experience is an important factor in understanding a problem. A computer may be capable of carrying out evaluation of success criteria for a design, but it is not capable of reasoning about what criteria are appropriate in the context of a given design problem. Proponents of artificial intelligence claim the situation can be changed but have brought forth no convincing evidence to support the assertion. The very fact that artificial intelligence is artificial eliminates the influence of commitments and personal risk.

7.3. Design Tool Leverage

The question of computer aids to engineering design can be viewed from a different perspective by considering the nature of a tool. In this thesis, care has been taken to emphasize use of the expression "design tool", reflecting the idea that a tool supporting design decisions need not reside in a computer to be useful. Conversely, use of a computer does not of itself assure the design tool will be an improvement.

The purpose of a tool is to provide an extension of ability where the human is limited. Thus a wrench provides grip and leverage as an extension of the hand. A design tool, then, should provide a similar extension of mental abilities to help the engineer be more effective at design.

Rather than automation of design, tool is a more appropriate expression reflecting the definition of a tool as an extension, not a replacement, of human abilities. The author prefers a goal of building design tools which extend the decision-making abilities of human engineers, rather than replacing engineers with machine decision-making. For such a tool, three design goals can be identified: integration, checkable results, and initiative left to the user.

First, the system must cooperate with integration of the design task. Specifically, those aspects which require a large effort to establish, such as the geometry of a drawing, should not require re-entry in a different form to access a different kind of tool. Vertical integration of the process of dealing with geometry suggests a system which, for early decisions, supports sketches, but allows that same sketch to be modified into a formal drawing at the point when a more formal, structured approach is appropriate. This same geometry should be transferable to the calculation methods used to verify the validity of the design.

As discussed in the sixth chapter, vertical integration is a greater need, and a greater challenge. As an example of vertical integration, consider manufacturability in material selection. For preliminary decisions, the approach given in the fourth chapter may be workable, but that broad approach is not as suitable later, when the problem must be reconsidered in greater detail. There are geometric considerations in manufacturability not considered at all in the fourth chapter (although they may be amenable to similar generalization). Knowledge must be integrated vertically - many versions interconnected to incorporate additional detail as the project proceeds to a more detailed description of the design. The level of knowledge required for design decisions will change as the level of the design moves further into details. One approach is to define several levels of detail in relatively separate layers and to record information accordingly. To preserve coherence, it would be necessary to build extensive connections between the contents of the various layers.

Another approach is to build a database access system which deals with a store of detailed information, but which is capable abstracting that information to the levels required by the progress of the design project. In the latter approach, what "intelligence" may be in the software is oriented toward providing the designer with information in a form to support decision-making, rather than on trying to make a decision.

Design tools must integrate with the user and with each other. A design assistance system is not necessarily a single tool, and need not, in fact, be resident in a single kind of system. Rather the designer needs a collection of tools, each tool suitable to particular types of problems. The designer's tool box is an analog of the mechanic's or carpenter's tool boxes. Each tool has uses which must be clear to the user, allowing ready adaptation to the particular problem context. Each tool must be workable in concert with other tools, since they will interact as they are used. However, it is not necessary that these tools all operate within the same, computer-based environment. In most cases the engineer can act the part of an intelligent interface to connect together the services provided by different kinds of tool. This requires, however, that each component of the design system is easily integrated with the human user.

Not every aspect of the problem will require the same treatment. This includes both the type and depth of treatment. No one tool may be a prerequisite to others working correctly. In essence, this restates the knowledge attribute of flexibility in terms of design tools. Since the point of entry into, and path through, a design is unknown, no entry point may be selected in advance by a design tool developer.

Second, the operation of each tool should be understandable and checkable. In some cases, such as the program for verifying flywheel designs discussed earlier in this chapter, the checking is done in advance through extensive testing. Such a tool may be thought of as having been certified by its authors. An important caveat of such certification is a statement of the limits; one must not unknowingly allow such a tool to extrapolate beyond its validating tests. The company tested and correlated flywheels bracketing all the concerns expected to be reasonable for design applications, to assure the tool would remain reasonable in use. Even so, there are some types of design feature which were excluded, and when these features are used, additional physical testing is necessary.

For more general-purpose tools, certification in advance is not feasible. In this case, each application must be checkable, by comparison to another evaluation method or by monitoring the operation of the tool itself.

The simplest form is for the engineer to have completed an analytical evaluation by hand, to obtain a working solution prior to using the more complex tool. This is a necessary practice, particularly for the complex software used for many analyses. Another potential form of solution is sample problems which incorporate the features of the system to be exercised into examples which resemble intended applications. In most cases an infeasibly large number of test cases would be necessary. Another approach is to slowly escalate the analysis, adding greater detail in geometry and loading in small steps, after the prior steps have been verified. The top-down, cumulative approach is also often a minimum-commitment approach. Checkability may be thought of as the knowledge attribute coherence. A simple solution to the problem, which the engineer can understand and verify, must be connected to the more detailed solution produced by the tool used for analysis.

Third, the system must leave the initiative with the user. Tools are intended to serve the user's needs, not the reverse. Price (1985) observes when the computer takes control in an automation setting, the human operator is unable to monitor what's going on: "automation can starve cognition." The question of initiative extends to how the situation is structured. An evaluation technique must be organized in a way that allows its use in many different kinds of situation. A design tool must be organized in a way that permits the user to direct its operations.

The discussion of knowledge structure in the sixth chapter incorporates an important aspect of initiative. The structure associated with a solution approach must be imposed by the engineer in the context of the particular problem to be solved. Therefore tools to be used to assist in solving problems most not impose a pre-selected structure. As an extension, if structure seems unavoidable in a solution method, the developer must strive to have that structure match reasonably with the structures an engineer might impose during design.

Compare the versions of fracture mechanics knowledge given by Liebowitz (1969) and Marriott (1988). The authors compiled by Liebowitz have imposed a structure of mathematical rigor which limits the broad applications of the information in their articles. This is indicated, for instance, in the geometries of the cracks and structures considered. The solution methods are constrained by the standardized mathematical approach, and only certain limited, regular geometries are amenable to convenient solution. These geometries are adaptable to a variety of practical engineering problems, but how well they match the needs of a design situation is left entirely to the judgement of the individual engineer and is not addressed by the authors of the articles. In contrast, Marriott's development is driven from problems of engineering interest, rather than from methods of solution. Marriott explicitly addresses the breadth of validity of his methods, supporting the engineer in making a judgement of whether the solution will be valid. The comparison between the approaches of Liebowitz and Marriott illustrates how both researchers and design tool developers can limit the engineer's initiative in design. The way knowledge is organized and presented is a major influence on its usefulness in design.

The discussion of expert systems, earlier in this chapter, indicates the most obvious mode by which initiative can be taken from the user, when the system forces the decision. But other tools also limit flexibility, and thereby limit initiative. Detailed numerical analyses, for instance, enforce a complex, details-oriented view of the design, to the detriment of the broad overview required to coordinate the effects of many disciplines. Use of simple, easily-implemented hand calculations, which was discussed by several engineers interviewed by Comer (1987), are important aspects of engineering practice which are largely neglected by research.

As an example, consider the use of the large, commercially available finite element analysis systems. The software provides so many features that effective use of the tool requires a dedicated expert. Several of the engineers interviewed by Comer (1987) were employed full-time at operating these programs and interpreting the results. Finite element programs have been built in a way that emphasizes the modeling technique, the behavior of the tool, over the problem to be solved. A design goal for decision support is to assure the tool allows the user to concentrate on the problem, rather than on the solution method.

Design proceeds by a connected series of reasoned judgments on the part of the engineer. If the initiative is given to a machine in this process, the chain of judgement is interfered with, which limits the effectiveness of the decisions instead of enhancing them. This point is also made by Cooley (1986), who discussed loss of skill and design judgement among engineers and architects in a design environment heavily dominated by use of computer aids, in analogy to loss of skill among manufacturing workers due to factory automation. Cooley feels that designers working in a computer-dominated environment lose skill because they are prevented from dealing with complete problems. Flexibility in application is needed to leave the initiative with the user.

The author has reviewed above the limits of design tools and recommends some standards for judging improvements to them. These standards are:

tools must deal with knowledge for design in the context of the problem being solved, which means the knowledge being manipulated must be coherent and capable of being integrated into the problem;

the results produced by tools must be checkable; and

the tools must leave initiative in decision-making to the engineer.

7.4. Decision Support

The previous two sections argued against automation of design tasks. How then is a tool to be used to help an engineer make design decisions? The author will answer that question in this section by examining what humans and computers are good at, and the forms of help an engineer uses currently. The key is to think of tools in terms of leverage, supplementing rather than supplanting engineering skill. Much of what is known about interaction between human and computer in decision making deals with process control, such as in chemical plants, power plants, and other production processes. The lessons learned in the study of process control are largely adaptable to the study of design decisions.

There are demonstrable limits to human abilities. For instance, Moray (1986) lists thirty-four limits on effective human decisions. Weiner (1985) classifies these limits in two areas: manual control and mental calculation. His view, which emphasizes keeping the operator "in the loop", is to "unburden operators from details-oriented, routine, repetitive, instantaneous control, and emphasize the human as big-picture strategist." This thinking is reflected in efforts in some design areas to construct and use high-level design languages, such as for integrated circuits (Begg, 1984). An approach based on high-level, top-down design relies on the assumption the problem area can be decomposed hierarchically, that is, each unit can be treated in relative isolation from others at the same level of detail. Software design, for instance, is approached in many cases with the intent of deliberately creating such isolation as a means to preserve reliability and maintainability and prevent unintended effects (De Marco, 1979).

There are many design problems, however, which are not amenable to treatment in isolation. The failure prevention issues discussed in the fifth chapter are heavily influenced by interactions between separate aspects of the design. One reason that manufacturability is now a problem is the past practice of making a high-level separation between design and manufacturing planning, commonly referred to as "over-the-wall" design. Typically in mechanical design problems, the engineer cannot simply specify functions and have a supporting design system provide the details. The engineer needs to remain in control of the design process, from the initial, global decisions down to the detailed level. Similar vertical integration of control for process operators is argued by Moray (1986).

Winograd and Flores (1986) argue that a decision is not a decision without commitment to action. A person has not made a decision unless they have the means to take action and implement it, nor would one take seriously a decision from a person who did not appear able to follow through on that commitment. By this definition, computers are incapable of making any but the most trivial decisions. Computers are capable, however, of enabling access to information through storage and communication facilities. Engineers in design do not want tools which make decisions because they don't want decisions imposed on them. Engineers do want tools which make decision-making easier and more effective by making access to appropriate information easier and more effective. But unlimited access to a vast store of detailed information, it has already been observed, is a step in the wrong direction. The data made available must be integrated vertically, allowing access not only to information on the problem, but also to information at the level of abstraction appropriate the engineer's current needs.

Woods (1986) points out two differing views of the consultant. One may think of a consultant as someone who is called in to take over and solve the problem, on the assumption the original people involved couldn't deal with the problem and are willing to abdicate their authority. In contrast, one may view the consultant as a data source, providing an information management service for the decision-maker. In the second view lies the basic philosophy for the development of decision support systems.

A consultant who enters the scene in the middle of a problem and supplies a single decision is undesirable for a number of reasons. Woods (1986) points out that decision-makers (humans or systems) which are good at particular subproblems are not necessarily effective in the context of a complete problem. This points again to the importance of problem context. More importantly, the designers cannot allow a portion of the design to remain outside of their realm of control and understanding. Reasons for this requirement include the context of the design, how the specialist solution connects with or influences the remainder of the designed system. The engineer is also concerned about professional ethics and professional liability and will take the attitude that "I'm responsible for this system, the expert can just walk away if something goes wrong." Moray (1986) sets two design axioms for process control systems: the actor with ultimate authority and responsibility for the decisions must have complete control over the process, and must have complete access to relevant information. These axioms provide important guidance to design tool developers as well.

Coombs and Alty (1984) describe three roles which human experts play: solution-generators, tutors, and consultants. The first two are most easily automated, yet are the least useful. The expert system as consulting information source and manager, which they call a "knowledge broker", is both the most sought-after tool, and the most difficult to build. It is not yet clear how other tools might be built to serve the same function. There does seem to be a large area in which tools could be built to enhance the human engineer's ability to broker knowledge for himself.

The tools developed and discussed in the fourth and fifth chapters indicated some portions of the information management function can be served by tools based on technology which is already available. In the fourth chapter, the necessary manipulation of manufacturing knowledge was done prior to construction of the tool. It could be argued that the resulting compact knowledge obtained was itself a tool for design, and required no additional organization. No computer support is necessary for effective consideration of manufacturability using the methods of Chapter 4, although the author built a simple database of metals which made it easier to apply the manufacturability principles in the context of other, equally important material selection issues. In the fifth chapter, several approaches to simplifying access to a large and disorganized collection of information about material failures were explored. The design checking methods give an example of the goal of leaving all the decision-making to the engineer. The case studies retrieval system, and the use of "one-liners" do nothing but provide the designer with cues about potential problems. Even the decision to consider the problem at all is left to the engineer.

A vertically-integrated, flexible and coherent design database can be implemented in larger systems. Schank (1984) describes a program which was originally built as a natural-language processor, but which grew into a general-purpose data query system with a limited vocabulary. The system is capable not only of obtaining information from several different data stores, but also can answer some questions about that data, such as high and low values, trends, bounds, and connections between entries. Notably, the system is capable of generating abstractions from the data retrieved. This provides an example of a system built with the use of artificial intelligence technology, but which leaves the intelligent decision-making to the human user.

Woods (1986) makes reference to goal-directed knowledge representations. Fitting engineering knowledge to its intended user, the engineering designer, has emerged in this thesis as an area where fruitful engineering research could be undertaken. Instead of arbitrary collections of information, tools must be built and knowledge collected with an appreciation of how they will be used by the human decision-maker. In the sixth chapter a "fitness-for-purpose" criterion was described which could be applied to both knowledge for design and design tools.

7.5. Design of Design Tools

The preceding discussion has given two kinds of guidance about design tools. First, two types of design requirements have been identified for which the support of design tools are not generally available: pre-detail requirements, and quality concerns. Second, three design requirements for design tools have been stated: integration, checkability, and user initiative. In this section the author will discuss a relatively new type of CAE system, which has been developed especially for use early in mechanical design problems. Following that will be a more general discussion of further improvements which could be obtained in tools to be developed.

The requirements for design tools may be explored by considering the working environment provided by spreadsheet software. All spreadsheets provide such facilities as sorting, searching, algebraic and logical formulas, and cross-references to other cells, allowing a spreadsheet to act as a general-purpose basis for numerical databases, tabular calculations, and calculations to support graphs. Many are capable of keeping text as well as numerical data in a cell, providing column and row headings; a few are capable of manipulating text in much the same way as numbers are handled. This provides most of the capabilities required to support engineering calculations.

The principal advantages a spreadsheet type of environment offers are ease of use and the way it leaves initiative with the user. User-friendliness in a spreadsheet does not extend to leading the user through the problem. On the contrary, the system itself is entirely unconcerned with the subject matter of the model. Selecting a suitable model and implementing it is left entirely to the user; the system makes the job of modeling less difficult but will do an equal job in all respects whether the model itself is good or bad.

A spreadsheet environment leaves the initiative with the user. The model is checkable for two reasons: users must build the model themselves since the spreadsheet does not supply the model, and the intermediate values of the solution are also stored in spreadsheet cells, where they can be inspected.

A spreadsheet environment can support integration because the model is independent of the software itself. In fact, several different models can be built, using different regions within the spreadsheet. This allows the user to build a connected collection of models. Vertical integration can also be obtained by using the common feature of links between spreadsheet files, which allow values from one file to be extracted for use in a second. Thus a file of detailed material properties may include maxima, minima, and mean values for that material; another, higher-level spreadsheet would extract those overview values from several spreadsheets, allowing a summary of the properties of many material classes to be brought together.

As an example of vertical integration, consider the design and manufacture of a molded plastic part. The engineer working on the design originally develops a rough idea of the class of material to be used, based on such questions as the properties to be obtained in the final part, cost, and formability. The engineer could use a spreadsheet which contains summary information about classes of plastic materials. This system is linked to other spreadsheets which contain more detailed information. In the top-level database, information summarizing these issues for each general class of plastic provides a reasonably compact point of entry into the selection of a material. One level further into the details, the same information for each of the varieties of a single class of plastic is provided in a second, connected database. The engineer refers to data at this level when a class meeting the needs of the design is being considered in more detail. This database, in turn, relies on summaries drawn from detailed spreadsheets giving properties for specific resins within a single plastic type. This, lowest level version of the database is entered to finalize the details of the material and its processing. Adding a new resin at the lowest level and recompiling makes that resin's unique properties available throughout all of the levels of the system. This hierarchical database concept was developed with the author's participation for a plastics molding firm; a similar database was constructed by the author for personal use for steel, aluminum, and copper alloys.

Applying the hierarchical approach to material selection for molded plastic parts provides some measure of economy and minimum commitment in the material selection process. It also leaves much of the decision to the user. For instance, formability in molded plastics is a complex issue, dealing not only with flow temperatures and pressures for the material, but also with die life, molding cycle times, and other aspects of the manufacturing steps. To be able to compare different plastics for formability, users must determine appropriate figure-of-merit formulas which indicate their own needs and priorities in the context of each project.

While a spreadsheet environment is capable of supporting a wide range of performance and quality evaluations, it is limited to those issues which can be expressed numerically. There are tasks for which the spreadsheet is not well-suited, both numerical and qualitative. Thus a spreadsheet is not the only tool which an engineer needs. It does, however, provide a useful illustration of the kind of environment needed for the tool to provide engineers leverage in design.

CAE systems have been introduced to the market, each specifically marketed to address early design by mechanical engineers. They use the notion of variational geometry, in which the dimensions of the picture being presented is dependent on the outcome of calculations being performed. They include the ConceptStation (Aries, 1990), I-DEAS (Systems Dynamics Research Corporation, 1990), and Mechanical Advantage (Cognition, 1987) workstation software packages as well as less expensive software based on the variational geometry approach, such as Cedar (MCAE Technology, 1990), Analytix (Saltire Software, 1990,) Mechanical Engineers Workbench (Algor, 1990), and Design View (Computervision, 1991). These systems seem to offer an improvement in computer aids to engineering design, and offer examples of implementation of some of the design goals given above. Mechanical Advantage comes closer to these goals, and will be the main subject of the discussion.

The Aries system uses solid modeling as the basic graphic form, on the assumption that engineers developing a design will reason about the part in its true, three dimensional form. The main analysis method relies on finite element modeling. The pre-processor to the finite element program has some distinctive features which represent improvements over most approaches. For example, boundary conditions (loadings and constraints) are applied by the engineer directly to the solid body. The mesh generator subdivides boundary conditions as necessary to suit the individual elements, without the need for the user's intervention. This approach allows the engineer to concentrate more on the loads and stresses in the solid, less on the modeling technique itself.

The author has approached solid modeling as a design method with suspicion. The author has experimented with the solid modeling capabilities of the Aries software. The sales argument is that engineers think about individual objects as three-dimensional entities. When developing a concept of a layout, the engineer writing this thesis doesn't think in terms of individual objects, rather he thinks in terms of assemblies, visualized in terms of one to three principal sections through the assembly. A two-dimensional visualization method - whether that is on paper or on a video screen - can be appropriate.

Solid modeling has a number of uses, however. The author has found solid modeling to be very useful for putting together an assembly which does not then act as a moving mechanism, for instance an electronics package. In fact, fitting a number of objects into a defined space is one application where solid modeling seems to be quite effective. In addition, solid modeling is very effective for developing and presenting to a prospective client a concept of a new system.

The Mechanical Advantage (Cognition, 1987) relies on two-dimensional representations, and in many ways the operation of the design system resembles a spreadsheet. The user sets up formulas describing the solution, linking certain variables to the dimensions on the drawing, others to related considerations such as the properties of the material chosen. Some of the variables may be fixed, others will then be determined by evaluation of the formulas. This provides horizontal integration between the picture of the design and the relevant calculations.

These CAE systems represents an improvement in the partnership between engineer and computer. The human is better at deciding what evaluation method to apply, therefore the user puts the formulas in place and sets parameters. The computer is better at running through the mathematics and maintaining the graphics presentation, and those jobs are left to the computer under the directions set by the user.

The computer's modeling ability does not deal with the judgments required in developing a design. The computer's function is to structure data in a way that makes human judgments easier. If the user elects to use a poor evaluation approach, the computer, acting simply as a tool, carries out that evaluation. Initiative in use and responsibility for results is left entirely with the user. The Cognition sales representative explained, "The Mechanical Advantage can make a good engineer better at design, but it can't make a bad engineer a good one." A sales representative for a solid modeling package commented, "graphics like this can make even a bad engineer look good - for a while." These comment are true for all design tools.

The system is checkable because the formulas solved by the system are expressed in a plain, mathematics format which allows the engineer to verify them rapidly. Once a set of formulas has been given to the computer, if checking is required a test problem can be easily constructed.

Software of the type just discussed improves the connection between pre-detail and detail issues. Most systems can transfer the final sketch to an ordinary CAD system for further, detailed development. As far as they go, this kind of workstation represents a definite improvement in computer aids to design. However, these systems are still concerned primarily with geometry and performance considerations. This leaves quality goals, and many pre-detail concerns, still unsupported.

In the fifth chapter, an extension to a system like the ones discussed above was discussed, which allows better consideration of basic quality concerns. Such an improvement would allow some failure prevention to be considered during early design checking, giving at least one concrete improvement which would make these tools still better.

The spreadsheet environment provides a useful analogy, defining by example design goals for design tools. Although the spreadsheet itself may not be a component of a design tool, the passive behavior of the tool is important.

Winston (1984) gives four goals for engineering workstations of the future. It is useful to close this section by considering Winston's four goals in the light of the ideas developed in this chapter.

First, Winston says "we must be able to talk to the machine in our own language". This should not be mistaken as the need for a natural language capacity in the system. An engineer is necessarily multilingual, being fluent in one or more human languages (spoken and written), in an extended technical vocabulary, in graphics and drawing symbols (which are another technical language), in mathematics, and most likely in computer operating systems and programming languages as well. The engineer is willing to talk to the machine in any combination of these. The Mechanical Advantage system, discussed above, provides an example of talking to the machine in a convenient language, which is a blend of graphics, mathematics, and programming language.

Second, Winston feels engineers must be allowed to work the way they want to, not controlled by the limitations of the machine. Surely any design assistance system will have limits; the trick is to assure those limits do not unduly constrain the way the tool is used. Two such constraints are the question of initiative, which has been discussed above, and access to an integrated knowledge base, which was considered in the sixth chapter and earlier in this one. To be sure the machine's limits do not interfere with the user, three things are necessary. The system must give the user initiative and responsibility, as already discussed in this chapter. The system must provide adequate flexibility in its structures, to assure those structures will be reasonably adaptable to the problem. Flexibility was discussed in the sixth chapter, and applied in examples in the fourth and fifth. The system's structures must be reasonably familiar to the user to assure accessibility, an echo of Winston's first requirement.

Third, Winston wishes the computer to constitute a complete environment, with access to all tools, records, and references through the workstation. This is not a necessary constraint, nor is it a desirable design goal. In the author's own experience, there are advantages to using information from a mixture of media, such as reference books, calculator or slide rule, file folders of data and sketches from the current problem, and design files showing prior solutions, as well as the computer workstation. Limiting access to all reference sources through a single workstation can become a juggling problem, trying to keep all the information which must be correlated together on the terminal screen.

Integration of design tools into a working environment should be concerned less with the tools dealing with each other and more with the user's ability to use tools in concert. Therefore there is no particular need for the tools to all rely on the same workstation. Some tools are best provided in the context of a computer system, others are equally well-suited to other formats. There are many situations where the most effective solution to coordinating the work of several tools is to rely on the human engineer, who is after all the only truly intelligent interface available.

Fourth, Winston requires that "the hardware must be muscular and the graphics excellent". Certainly the user does not wish to be unduly delayed by a slow-acting computer, but the author has made the point that there are problems which sheer computational power cannot solve. This is true even when the computational power deals with manipulating symbols instead of numbers. An engineer does not deal exclusively with the workstation, and the cost of the system is a consideration.

Although excellent graphics are desirable, one must be careful to give proper meaning to that expression. As discussed earlier in this chapter, there is no particular value to highly detailed graphics in many design situations, in fact there can be disadvantages. However, the Mechanical Advantage workstation gives another version of excellent graphics, with the display of a sketch linked to a set of calculations in a well-integrated environment.

Winston's assessment leaves out the question of checkability, which is vital if the engineer is to remain responsible for the action of the tool. He addresses integration only indirectly. Winston is a computer scientist building models of users' needs based primarily on available technology and on how he would like to see computer technology advanced, rather than on the users themselves and how they work. For further development of design tools to be effective, that development must be based on a realistic understanding of how engineers actually do design, and what forms of support they need.

7.6. Conclusions and Recommendations

In this chapter the author draws four conclusions about how design tools may be better designed in the future.

1. Design tool development seems to be driven by what can be achieved by available computer technology. There are at least two reasons this is not appropriate. First, it is not correct to assume the computer is a requisite part of methods for improving design productivity. Second, although technology can limit tool development, tools must be developed primarily based on the needs of the tool user, rather than on what a particular technology has made available. There is a need for research which conveys to tool developers what needs must be met.

2. Systems to support design must take the form of tools to support the management of evidence for human decision-making. This provides the engineer with the necessary leverage to further improve the effectiveness of decisions. Automation of individual design subtasks will not, in the long term, improve engineering productivity because it will add even more isolation to subproblems which most be solved in relation to one another.

3. Design tool development has historically emphasized details-oriented, performance-oriented analyses. There is an equal need for tools which support evaluation of design quality issues, including manufacturability and failure prevention. There is an equal need for tools which support evaluation of issues prior to full definition of the design details.

4. Three design requirements can be stated as goals for design tool developers:

integration,
checkability, and

user initiative

These all contribute to a single rule of guidance for design tool developers:

> The developer must not presume to dictate how the tool will be used in design.

This single rule incorporates the need for flexibility and coherence, for integration, for checkability, and for the user to retain the initiative in operation. It provides guidance to both developers and purchasers of design tools. Both Aries and Cognition have claimed to have a majority of mechanical engineers in their development staffs. Even assuming those staffs have design experience, which is doubtful, this qualification does not allow them to anticipate the attributes of other design projects. Without design experience, developers must be doubly careful not to unintentionally constrain the uses of the tool.

Chapter 8
Design Philosophy

Like all human activities, engineering design has an underlying philosophy. The philosophy behind an approach to design helps to define how to approach problems, and expectations about the form the solutions may take. There are many flavors of the philosophy about engineering, and therefore many kinds of assumptions used as starting points for design problem-solving.

There is no clear right or wrong approach to design. However, the author's practical experience as well as research on the subject indicates there are more and less productive avenues of thought. In this chapter the author discusses a few aspects of design philosophy, some of them productive and some of them counterproductive. The author finds the occasional calls for "rational design" (that is, design based on rigorous solutions methods) very troubling, and the concept will be discussed in the first section. Engineering education and research are influenced by, and influence, engineering design. In the second section areas of design research appropriate for consideration in the future will be discussed. In the third section, the author considers the relationship between design and engineering education. There is a strong connection between design research and design education, not only because engineering students will later become engineers doing design, but also because engineering educators are also researchers who produce information which engineers are expected to apply.

8.1. Rational Design

Psychologists and philosophers have rediscovered the debate about human rationality in the past several years. The question, originally asked by the philosophers of ancient Greece, is whether a human will act in a rational, predictable manner, or whether human actions are essentially unpredictable.

Rather than deal with the debate in great detail, the author will examine the meaning of rationality in the context of the development of tools and methods to assist in engineering design. Rather than attempting to decide if man is rational, it is only necessary here to consider the influence of particular views of rationality on how engineers make decisions. In particular, the meaning attached to the word "rational" must be carefully considered.

It must be assumed that an engineer will act rationally. To assume otherwise is to embrace a whole range of irrational, even mystical and metaphysical approaches to design which are difficult to defend and would probably not be as successful. But especially in design there is a distinction between a rational act and one derivable entirely from strict logic and rigorous mathematics.

Winograd and Flores (1986) make a distinction between human rationality and the constrained version they call "the rationalistic tradition", which might also be called the narrow view of rationality. The rationalistic tradition reduces things to fixed procedures, require problems to be solved in the right order with each step properly completed before proceeding, and intermediate values taking the correct form. This focus on only certain narrow aspects of human rationality often leads to attitudes and acts which are not really rational at all, when viewed on a broader perspective. A procedural, sequential approach to problem-solving, for instance, ignores any possible limits on time, or on the capacity of the engineer to understand the intermediate values. Narrow rationalism leads to ineffective use of engineering resources by focusing only on one aspect of the problem.

The narrow view is also a problem in mathematics. Van Dormolen (1986) comments that a formalist considers only formal mathematics to constitute mathematics. Polya (1957), a great informal mathematician, felt the rigid, formal, Euclidean proof in mathematics has limited value for at least two reasons. A proof may demonstrate, through logic, the validity of the assertion but it does not reveal the methods used to discover the principle. Polya considered a Euclidean proof to be one of the least effective ways to help students understand a concept because the proof does not necessarily convey how the principle was discovered, what it means, nor how it can be applied. He felt the means by which a concept was discovered was important information, often deliberately ignored because those aspects of the process lacked rigor.

The philosophical question of rationality has direct influence on engineering. The last few decades have seen a series of authors proposing to put rationality into design. Weck (1966) proposed putting the codes and standards governing the design of weldments on a more rational basis. It is now possible to evaluate the stresses and material properties through a weld of any geometry, he argued, rather than rely on simple rules of thumb and safety factors. Weck would have had each individual weld evaluated on its own merits through detailed calculations. There are several flaws with this type of approach. It ignores the economic realities associated with allocating engineering resources to design of a system which incorporates perhaps thousands of welds.

Weck seems to assume from the simple structure of the rules of thumb that their origins are simple, ignoring the wide scope of both practical and theoretical work necessary before such a rule is considered valid for use. By rejecting factors of safety, Weck assumes that it is possible to account for every conceivable event, which is clearly unreasonable. The ability to calculate stress distributions in great detail has no direct relationship to the ability to evaluate the adequacy of a structure under a load; in other words, the rules of thumb would simply have been moved to another location in the reasoning process. Weck's "improvement" to the design of weldments, had it been implemented, would have represented a large step backward for engineering by adding to both cost of design and confusion during design. This example illustrates how an improvement which appears, in the narrow view, to be rational, can be irrational when viewed from a broader perspective.

The rationalistic tradition is predicated on the assumption the world is orderly, and imposes orderliness on engineering acts intended to deal with that world. But the world is not orderly. Rittell and Webber (1972) use the word "wicked" to describe design problems in architecture and urban planning. The world is not static, it is constantly changing. Attempts to impose order are counterproductive because they are attempts to make the world stop changing. Exact calculations have value in some situations, but effective use of engineering resources requires developing other, less precise calculations for design.

Every calculation, even an "exact" one, is inexact. Mathematical rigor is defined by arbitrary, global limits on what kinds of assumptions will be allowed. But assumptions are only justifiable by verifying reasonableness locally, in the context of a particular problem. An irrational (arbitrary) set of rules are the basis for "rational" design.

Engineers have long recognized that logic and mathematics are useful tools, but as tools they cannot dictate the conditions for their own use. Alexander (1968) observes that the choice of a particular logic is arbitrary, and "logical methods, at best, rearrange the way in which personal bias is to be introduced into a problem."

One ordinarily assumes that scientific and engineering research are passionless, fully rational activities. A counterexample was related by St. Denis (1989), who published in 1950 the first paper which analytically evaluated seakeeping, the motions of ships under the influence of ocean waves. At that time, and for a few years afterward, St. Denis was repeatedly told by naval architects and naval officers that seakeeping was not a subject worthy of research. But in 1955, four of the nine papers presented at the annual meeting of the Society of Naval Architects and Marine Engineers were on the subject of seakeeping. St. Denis was pleased to note that the attention had shifted from criticism of the need for his research to critical evaluation of his work's content. This episode illustrates that fashion and personal opinion, not just rational reasoning, are important factors influencing research.

There is no such thing as the "right" answer to a design problem, there is only a distinction between one solution better by some measure than another. Even a relative definition depends on what aspects are chosen for emphasis in the decision just as the outcome of an optimization exercise depends on the choice of an objective function.

8.2. Design Theory and Design Research

In this section, the author will examine the form design theory might take, and consider the potential of mathematical knowledge as a model which design theory might follow. Following that, the discussion will turn to what areas should be considered in a research program which is intended to add to design theory or support the design activity.

Over the last fifty years, engineering has progressively been moved to emphasis on science. More specifically engineering has moved to reliance on mathematical expression of scientific principles. It is believed that mathematics and science provide the only principles upon which engineering proceeds through its work (Gregory and Turner, 1972). Rather than debate the issue, this section examines the nature of the concepts that are the basis of mathematical thought and applies them to engineering design.

Design theory must not be limited to analogs of mathematical theorems, which are simply descriptions and proofs in detail of certain notable relationships. Mathematics can also be considered in terms of how mathematical problems are approached and solved. Polya (1957) dealt specifically with problem-solving. Recently, mathematics educators have readdressed the question of problem-solving (see, for instance, Ballem, 1983; Dorfler and McLowe, 1986; or Kantowski, 1981); their experience provides some useful guidance for development of a comprehensive design theory.

Van Dormolen (1986) identifies several important aspects of mathematics, all of which work together to define how problem-solving can proceed on a mathematical basis. These include:

- Theorems: formalized and prescriptive, and dealing with those limited particulars which are formalizable;

- Algorithms: simple procedures describing how to accomplish the solution to certain classes of well-defined problem;
- Logic: defining allowed and disallowed methods of making judgments;
- Methodology: strategies and heuristics dealing with approaches to problems; and
- Communications conventions: standardized methods for conveying information about problems and solutions.

These can be adopted outright as a description of the key aspects we should seek to develop for design theory. Nadler (1981) gives three general classes of knowledge required for use during design:

- Knowledge about design, concepts guiding problem-solving;
- Knowledge in design, organizing principles for dealing with a problem; and
- Knowledge for design, problem-specific information which is called upon as the problem-solver recognizes deals with specific aspects of a solution.

Knowledge for design may be likened to theorems in mathematics. Knowledge for design appears to be the only focus of research in engineering at present. Knowledge in design deals with logic, algorithms, and communications. The author has emphasized the importance of communication for design. Knowledge about design deals with methodology. Note that methodology is not constrained, specific algorithms or recipes for action, but rather is broad approaches to problem-solving.

A complete theory of design, encompassing all these areas, would be a large and complex work. It is possible, in fact desirable, to develop independent portions of this theory, provided a developed segment has within in a context showing how it fits the whole of design theory. Design theory, like design tools, need not exist entirely within a single system. A series of separate theories, each limited to certain situations, is an entirely appropriate interim solution. In fact, multiple approaches to a problem are desirable and useful (Olsen and Anderson, 1986). Asimow (1962) mentions that art and architecture have always had a variety of schools or approaches to design at a time, and wonders why engineering has not. It is up to the engineer to choose that portion of theory which will serve to guide reasoning through a problem, just as the engineer must choose the appropriate tool from a set of tools.

Tomiyama and Yoshikawa (1987) are working on a single, unifying, well-defined and structured theory of design. Design is, however, concerned with ill-structured problems (Rittell and Webber, 1973; Newell, 1969; Simon, 1973), and for this reason the author is unsure an approach predicated on finding structure in design can succeed. Much of design is not understood in an explicit way, limiting the scope of such a theory. The cognitive psychology of decision-making, for instance, is clearly important to an explicit understanding of how an engineer designs, yet theory in that area is still developing. This thesis is the first work the author is aware of which attempts to connect the more general psychology of problem solving with the specific problem-solving issues in engineering design.

Engineering design, like all forms of design, is first and foremost a pragmatic, practical activity, aimed at producing descriptions of objects of practical value. Asimow (1962) recognizes that the only valid test for design theory is whether it works in practice. Alexander (1984) observes that theory is a fine thing to consider at leisure, but when confronted with a problem requiring solution, engineers are thrown back onto what they know and what they can do with their knowledge to solve the problem. It is inappropriate to think of design theory in absolute terms. Rather it is necessary to rely on theory as an underlying structure of organizing concepts, some of which will be useful in a given problem, and some which will not. It is difficult to anticipate how a tool will be used, what features it will need, or how its behavior should be structured, until the tool has had extensive operating experience. Design theory, as a tool to serve the needs of both engineers doing design and researchers supporting engineering design, has the same limit, requiring extensive experience of use to become established.

While a theory providing prescriptive norms for design decision-making may be useful to guide design research, it is necessary to credit engineering designers with considerable capability at decision-making. Although design research is concerned with improving engineering performance, it must be noted that engineers are already quite good at design, but even marginal improvements are important in a competitive market. It is not clear that engineers will benefit from instruction in how they should change their problem-solving methods; the author believes there has been a net disbenefit from attempts at this kind of instruction in the past. Design researchers, however, will benefit from a clearer understanding of how design is actually accomplished. To use the terminology of cognitive psychology, design theory at the current stage in its development must be descriptive, rather than normative. It is not appropriate to tell an engineer how he must think. It is appropriate to tell researchers and design tool developers how engineers think, with the goal of improving the support the design process receives.

Researchers should be reluctant to suggest potential improvements in design thinking, doing so only when the researchers themselves have substantial design experience. Researchers should be eager to develop their collective understanding of design as it is practiced daily by engineers.

It can even be argued, by analogy to mathematical knowledge, that design theory can never be a single, monolithic entity. Russel and Whitehead's ambitious *Principia Mathematica* was intended as a complete statement of all that was known about mathematics. Hofstadter (1979) discusses the assumption that such a complete statement of mathematical knowledge was possible, and concludes that *Principia*, although an important work, was never destined to achieve its author's goals because the goal was not achievable.

The author's opinion is that design theory must necessarily remain a collection of loosely-connected principles. This is based on application of Godel's Theorem of Incompleteness, originally developed for number theory but explored in its general implications by Hofstadter (1979). Put simply, any closed, formal, logical system is insufficient for all reasoning because it is incomplete. This incompleteness can take two forms. First, the closed system incorporates concepts which cannot be proven by relying solely on the other concepts contained within the system. For instance, Euclidean geometry must rely on a foundation of postulates which are assumed to be true. Second, concepts within the system can be combined in ways dealing primarily with things not contained in the system.

Applied to engineering design, Godel's Theorem can be stated this way:

Innovation cannot be obtained from procedures and precedent.

Design necessarily takes place at the edge of what is known, usually applying what is not fully understood.

Our focus on mathematics tends to be restricted to theorems, making incompleteness of the closed logical system difficult to perceive casually. But technique, that is, heuristics and strategies for manipulating theorems are equally important, both in mathematics (Polya, 1957) and in engineering (Nadler, 1981). When one includes such methods in the system, the incompleteness of even a formal mathematical system becomes more obvious at the intuitive level.

The development of design theory can follow mathematical methods, within limits. Four concerns have been discussed above. First, design theory must encompass not only what knowledge is used by engineers, but also how, why, and in what forms that knowledge is used. Second, design theorists need not strive for a single unifying conceptual system, since formulations of design theory which reach toward that goal will be unwieldy and it is likely that such a system is unattainable. Third, fragmentary theory of design is acceptable and useful in the short term, and may even be the most appropriate approach for the long term. Fourth and most important, design theory should not be developed with the intent it will be used to change the ways engineers approach problem-solving now.

Not only design research, but engineering research in general influences engineering productivity during design. Research is information-producing, and design is a knowledge-using activity. The needs of design must drive the ways knowledge is structured for use. Barrett (1981) observes that research results are not in general useful to designers. Figure 3 illustrated the conceptual distinction between a direct and an inverse problem. Most research products are developed in the context of, and for that reason best suited to deal with, direct problems. There is a need for researchers to develop knowledge into a form more amenable to design problems. A senior member of the research division of an aluminum company observed that the ability to deal with inverse problems is a sign of maturity in an area of knowledge. Well-established areas of engineering, such as kinematics, address the inverse problem explicitly. Such maturity is a useful goal for research in new fields, and researchers must begin to measure their success by the ability of engineers to use the knowledge in design situations.

A term used in the past for this issue is reduction to practice (Furman, 1970). The reduction may be done by the researchers producing knowledge, by others who would be the link between research and design, or by designers themselves (as seems to be the case at present). Knowledge structuring research is growing in importance as the flood of new information rises still further. Ideally, structuring to suit design application would take place as new information is produced. Approaches to developing the form of knowledge to suit design needs have been discussed in the preceding chapters. Further development of design-oriented knowledge engineering is going to be an important area of engineering research in the future.

Knowledge structuring is one aspect of a complete program in design research. The primary distinction between design research and other engineering research activities is the emphasis on method rather than on information about a single problem.

A complete design research program would incorporate work at several levels:

- Design, actually working on design problems and learning about design by doing it;
- Knowledge engineering, working on knowledge-structuring issues;
- Philosophy, thinking about broad methodological issues in design; and
- Applications, such as developing design tools.

These aspects of a design research program are interconnected. A research program dealing primarily with one aspect of design is then missing some vital context. Lack of a complete context is a contributing reason, for instance, for design tools that do not meet the expectations of their users (Freeman, 1984).

The research reported here is an example of such an intersection of subproblems. The author has been employed as a design engineer, working with structural and mechanism problems. Two knowledge-structuring issues have been explored in this research: failure prevention and design for manufacturability. Methods and tools to deal with these two areas were developed and tested. In the context of this experience, an underlying philosophy - an attitude toward both design research and design problem-solving - has been developed and described. Any beneficial effect on design theory as a result of this research has been due to the cross-products of these different categories much more than the individual portions of the work done.

Several authors comment that the primary block to improvements in development of design tools is a failure to fully understand design (for example, Gero *et. al.*, 1985; Tomiyama and Yoshikawa, 1985; Winograd and Flores, 1986, and Begg, 1984). The additional input needed to permit an

Flores, 1986, and Begg, 1984). The additional input needed to permit an advance beyond the status quo of design research is insight into how an understanding of design is applied in supporting design through tools and methods. Ernest Warman (quoted in Gero, 1985) makes a very useful distinction between "know-that", facts about a subject, and "know-how", knowledge about actions one might take. Although facts and information are being produced at a rapid rate by research, know-how is in danger of atrophy from lack of attention.

8.3. Design Education

Engineering is design. The author holds this opinion, and Petroski (1985) agrees. Design is the methodology and the attitude toward problem-solving, that makes the engineer distinct from a physicist working on similar problems. Therefore engineering education is design education. Since design is the essence of engineering, it should not be set aside in separate courses. To improve design education it is not necessary to have more courses in design, rather more design content in all the courses is required.

The distinction between direct problems and inverse problems in engineering (Dixon, 1966) was discussed in the third chapter, indicating that when a fixed, fully defined system requires evaluation, that is called a direct problem. An inverse problem refers to the need to define a system which meets certain desired parameters. At present, education emphasizes direct problems; design requires the ability to deal with inverse problems. This section will outline an approach to introducing greater emphasis on inverse problems.

Two observations form the basis of the author's opinion about design education. First is the need for the engineer to select the correct tool for the problem. Second is the fact that there is no single correct solution for most problems in engineering. These lead the author to recommendations about improvements in the way engineering education is undertaken.

A great deal of emphasis in the undergraduate curriculum is on developing knowledge about individual areas. The curriculum usually takes the form of a series of separate approaches to separate subproblems. In the course of developing students' knowledge it is also necessary to examine and teach the skill of choosing an approach, a tool, or a model which is appropriate to the situation.

Managers in industry comment that new graduates can make calculations, but cannot make judgments (Comer, 1987). This relates to the problem of tool selection. When in each academic exercise it is necessary only to manipulate a single model or method, the result obtained from that model is the goal and the student learns only to make calculations. By expanding the student's view to include making a judgement about a problem, the student is shifted to the position of selecting a suitable solution method, operating that method correctly, and drawing a conclusion from the result. The conclusion, and not the outcome of the calculation, must be the goal of the student's assignment.

A shift in emphasis from calculation to reasoned conclusion reduces the motivation to get "the right answer". With emphasis on drawing a valid conclusion, rather than making the calculation, students are free to approach the calculation - both the model to be applied and the particulars of its manipulation - in a manner suited to their own tastes and talents. The method of solution then becomes subordinate to reasoning about the situation.

Described another way, education can be improved by emphasizing products rather than procedures. The product is the engineer's purpose in industry, and can serve as a useful vehicle for a wide range of exercises. Naturally, it is necessary to learn about available procedures, but of equal importance is the need to learn about choosing a procedure to fit the situation.

Students will learn economy in problem-solving through experience. Students may have to choose between a clear path to the solution which involves considerable work and a new or more risky avenue which seems to involve less effort. The student learns to consider the cost of the evaluation when selecting a solution method.

An obvious objection to an approach based on individual reasoning is the difficulty of grading such a variety of solutions. It is not necessary to grade the problem-solving procedure in detail unless the conclusion seems ill-reasoned or incorrect. It is fairly easy to decide if a conclusion is well-supported by the evidence presented, and to quickly assess the validity of that evidence. Some aspects, such as economy of the solution, need not be graded at all since economy is its own reward. The author has used this approach when teaching a course on machine design, and grading effort did not seem significantly different from a more traditional approach to homework assignments.

There is a another, related area which Comer (1987) identified as needing improvement: the ability to communicate. Design-oriented assignments, which are evaluated on the student's entire reasoning approach, instead of the ability to manipulate a calculation method, require students to express evidence, evaluation and conclusion intelligently. Students who are better at communicating their ideas and thought processes will receive better grades for their work.

Van Dormolen (1986) comments that two levels of preparation are necessary for learning: not only learning the particular knowledge and its prerequisites but also learning about the conceptual structures of the knowledge, and approaches to its use. Stepping out of the evaluation and thinking about the process itself is equivalent to the artificial intelligence concept of metaknowledge, knowledge about the knowledge itself. Metaknowledge about engineering is engineering methodology and philosophy - attitudes and approaches to problem-solving. Significant improvements in engineering education can be achieved by addressing approaches to problem-solving in addition to specific methods. Design-oriented education enhances the student's grasp of methodology and philosophy which they will need for a career in engineering.

Chapter 9
Summary and Conclusion

The knowledge which is available to engineering designers, the forms used to record knowledge, and the modes of thought those forms allow have a strong influence on what engineers can accomplish in terms of developing and applying new ideas. In this thesis, research into the knowledge needed for design and the forms it should take has been used to develop an understanding of how tools can be built to better support design decision-making by an engineer. In addition, the nature of design and the thinking required of an engineer doing design have been clarified.

Engineers design by making decisions. Each decision requires evidence to support it, and to be most useful the form of the evidence must be appropriate to the form of the decision. Engineering productivity can be improved by improving access to evidence for engineers. Access to evidence can be improved by studying engineers doing design work and developing a better understanding of their needs, and by building design tools which specifically address the problem of finding and manipulating evidence to support engineering decisions.

9.1. Engineering Productivity

Improvements in manufacturing productivity have created a basis for hope that countries with expensive labor can still compete with countries where labor is less expensive. There is also a need, however, to deal with productivity during design - both improving the productivity of engineers and designing products for better productivity during manufacturing.

It is customary to think of productivity in terms of efficiency, that is, in terms of reducing the resources consumed. A great deal of current design research, such as design automation development, is based on a model of design productivity which emphasizes efficient decision-making. There is also, however, a need for development of more effective decisions, which may be thought of as improving the reliability of decisions: making fewer errors, anticipating errors instead of correcting them, and making decisions which have broadly favorable effects instead of locally favorable and globally unsatisfactory effects.

Consider, for example, the problem of economic fabrication of a design. The past history of extensive redesigns necessary to improve manufacturing efficiency show how product development decisions have been ineffective. Locally important effects, such as strength, have been allowed to dominate over other issues, such as manufacturability and maintainability, resulting in globally unsatisfactory designs.

Engineering productivity, and in particular design productivity, depends more on improving effectiveness than efficiency in decision-making. In the American automotive industry, competitiveness and productivity has been improved more by efforts to raise product quality (effective manufacturing) than by efforts to use fewer resources (efficient manufacturing).

Engineering productivity improvements in the future will rely primarily on making gains in assuring engineers are producing quality designs. Design quality incorporates product quality, but also includes issues one might not ordinarily relate to the end product, such as manufacturability. When decisions are made with attention to the quality of the design, those decisions will be effective, and will have global validity within the design.

The single improvement required to add to the engineer's ability to make globally effective decisions may be described as integration. By integrating the decisions to be made, and integrating the knowledge to support decision-making, the engineer can become capable of including a broader range of issues in decisions.

Integration cannot be achieved on the basis of details, however. During early product development decisions, which lay out the overall form of the product, the engineer is concerned with creating structure which serves as a basis for a quality design. Decisions at that time must focus on meeting a large number of limits, not only performance, but also reliability, manufacturability, and other concerns. Integration is necessary, but it is necessary to rely on rather general modes of knowledge to integrate between a large number of issues. Detailed information interferes with the concept development process by distracting from the broad issues which must be handled first.

Later, as the details are developed in design, a fully integrated knowledge base supports the engineer's focus on fitting particulars into the broad form already established. To assure that those details which will be problematic are suitably anticipated, it is necessary to have connections between the general and the particular forms of knowledge about an issue. To assure that details are created in keeping with the intent of the general form already developed, it is necessary to have connections between the particular and the general forms of knowledge about an issue. These connections within knowledge about a single discipline constitute vertical integration.

Vertical integration provides a basis for anticipating that certain details have sufficient importance that they must be considered earlier than would ordinarily seem appropriate. In contrast, current efforts at integration deal entirely with details, making no distinction between vital issues and deferrable decisions, nor between useful data and trivia.

Horizontal integration is the connections between disciplinary areas. These connections are most needed early in design, when emphasis in the definition of the product is to develop a complete structure of the product. It is necessary to recognize the cross-disciplinary issues and deal with them early in the product development process. Horizontal integration at the detailed level is possible, within some limits. Horizontal integration at a more general level will allow a broader set of connections to be accumulated. Horizontal integration at a more general level will improve the economy with which the connections between disciplinary areas may be handled.

A fully integrated knowledge base supports three activities during design. First, the interactions between different disciplines are accommodated. Second, reasoning at a high level of abstraction is supported to allow early decisions to be made without the hindrance of unnecessary detail. Third, detailed reasoning is supported and connected to the more abstract thinking which preceded detailed design.

Continuing improvement in engineering productivity depends heavily on developing means to improve the quality of the designs produced. Design quality in turn requires improving the effectiveness of design decisions.

Effective design decisions rely on a properly integrated knowledge base, and properly integrated support from the tools built to assist in design.

9.2. Minimal Knowledge for Design

The purpose of a single design decision is to advance the project by adding to the definition of the product. Design proceeds by a connected series of decisions, each one dependent on its predecessors. Each decision fixes a portion of the product's form, constraining subsequent decisions to those options which fit that form. Each decision makes a commitment which is expensive to retract or correct; the cost of correction is related to the time elapsed since the decision was first made.

It is to the engineer's advantage to avoid making unnecessary commitments during the design process, leaving details free to float until necessity pins them down. This is the strategy of minimum commitment.

How and when details are handled is an important aspect of a minimum-commitment approach to design. A detail fixes in place not only itself, but also its connections to the remainder of the product, and the effect of one detail commitment can be surprisingly extensive. A strategy for limiting detail commitments and their effects is to approach the problem from the top, with broad and general decisions first, then down into successively more detailed levels of decision. This is called "top-down design", and has been used successfully in many different disciplines, most notably software design. Each detailed decision is postponed until it absolutely must be made, achieving minimum commitment.

For a top-down, minimum-commitment approach to succeed, knowledge to support a succession of related decisions must be available. Early, general decisions deal with the same issues as later, detailed decisions, but within a broader and more abstract context. Therefore knowledge for dealing with the issues at several different levels of detail is necessary.

The knowledge required for a decision at a general level is more abstract than that for a detailed decision for at least three reasons. First, breadth of knowledge takes the place of depth, so early decisions can better deal with the scope of the product and the interactions within the systems which comprise the product. Second, if only detailed knowledge is available for a general decision, the engineer is required to abstract a solution at the level desired out of the volume of detailed information. This takes time and adds unnecessarily to the cost of decision-making. Third, excessive detail to support a decision is less economical, requiring greater thought on the part of the engineer, greater resources to maintain the supply of both knowledge and data, forcing decisions which make greater-than-minimum commitments, and requiring data which is usually expensive to obtain as a complete set.

To support minimum-commitment design decisions, minimal knowledge is required. Although it is not yet possible to prescribe a method for producing minimal knowledge, four attributes have been identified which can be used to evaluate how well a given knowledge representation meets the goal of minimal knowledge for design. These attributes are:

- flexibility,
- coherence,
- predictive value, and
- economy of use.

Flexibility in the knowledge base is achieved by recording the same knowledge through many different means, including qualitative, analytical, and quantitative types, and ranging in detail from the general to the particular. An effective mix incorporates narrative descriptions of processes and events, abstractions of general processes, numerical data, and analytic expressions which show the relationships between the data and potential applications. From this variety of forms, engineers can choose a version (which can include varying combinations of the available forms) meeting the immediate need. A number of differing versions provide knowledge with broad value in design settings.

Coherence is achieved by assuring that the different versions and forms which are combined in the knowledge base have adequate interconnections. Two forms of connection are important. The first form of connection is integration within the knowledge base, and is necessary to provide the means for knowledge about a design issue to be useful at several levels of detail as the design is developed. Engineers must be provided a visible relationship between the knowledge they are using for the current decision, the knowledge which was used at a more general level, and the knowledge which will be used in more detailed decisions to follow.

The second form of useful connection is that between the knowledge and the problems the engineer faces in design. The design problem sets a context, with constraints and problem types determined by the problem and the approach to the solution. To assure useful knowledge will be recognized and used in design, the knowledge which can be used to solve the problem must be recognizably connected to the context of the problem.

Prediction is the primary mode of reasoning during design. When the solution to a problem is incomplete and under development, it is not possible to engage in diagnosis or analysis. Not only the solutions to known problems, but also problems which might still arise must be predicted. If a knowledge representation is to be useful in design, it must provide a basis for predictions about the design. Predictions are necessary for both early, general decisions and later decisions dealing with local details.

Engineering is concerned with allocation of scarce resources. Engineers' time and engineers' cognitive capacity are scarce resources. The engineer in a design setting is equally concerned with developing an economical product, and with economically developing the product. Even if knowledge meets the other criteria given here, if it is not economical in use it will not be used. Economy is determined by the local context of the problem to be solved. For a general decision, excessive detail in the knowledge adds the expense of sorting and generalizing before the decision can be made. Flexibility, coherence, and predictive value are necessary for achieving a generally economical knowledge base.

9.3. Design Tools

The notion of design automation is a misleading one. There are both pragmatic and philosophical reasons it is not acceptable for human engineers to automate, and thereby relinquish control over, any aspect of the design process. Pragmatically, development of systems which generate designs have only been successful for applications in extremely well-defined and well-constrained settings, and operating in relative isolation from the remainder of product development. Methods to make an effective connection between the small subsets of potential solutions obtained from automation and the remaining, human-controlled design have yet to be demonstrated. The technology is not yet available to support any significant automation of design tasks, limited by both the capabilities of computers and the ergonomics of human-computer interaction.

Philosophically, it is not clear human engineers should ever be willing to relinquish authority or responsibility for product development. The engineer is responsible for the product, and a decision-maker with such responsibility must also have the authority to control the decision process. Engineers have no reason to allow portions of the design decisions to be made for them. Although there are aspects of this argument which deal with the ethics of human action, the most telling aspect deals with product liability. Regardless of the degree to which the design process is outside their control, individual engineers must decide whether they believe the product will be safe. Engineers, not a software product, will be held accountable if the product proves unsafe or otherwise unsatisfactory, indicating that both the engineer and society at large wish to retain human control of product development. The social climate is not yet available to support any significant automation of design tasks. There is no indication the social climate will change to favor design automation, nor is there any strong argument in favor of such a change.

Design tools are not, and should not be, used to make decisions for the engineer. Rather they are to assist the human engineer in making effective decisions. Design decisions require an extensive and complex supply of evidence. Design tools are used to organize, store, recall, manipulate and in some cases generate that evidence. Because the nature of the evidence to be used, and the organization to be imposed, varies from one decision to the next, the design tool must be flexible in the ways it can be used and responsive to the needs of the user.

The main purpose of a design tool is to help the engineer be more productive. One approach, reflected in the idea of design automation and reflecting an efficiency definition of productivity, is to improve productivity by automating portions of the design process, intending to reduce the human resources required to design a product. This is an analogy to the physical automation which has improved manufacturing productivity. Rather than systems which carry out design in isolation from human influence, and must therefore be checked, it is more appropriate to focus on systems and methods which extend human abilities, expanding the engineer's scope of action by providing leverage. Such an approach is based on a effectiveness definition of productivity, that is, improving productivity of engineers by supporting more effective decision-making.

Three design goals may be set for design tool developers. Like many design requirements, these goals can be stated in concrete terms, but the means to achieving them is not as clear. These goals are:

- integration,
- checkable results, and
- initiative left to the user

It seems customary among contemporary design tool researchers to create systems for particular, narrowly defined applications. What effort is underway at integrating aspects of product development, such as integrating CAD and CAM, focuses on the very detailed particulars of the issues to be integrated. A details-oriented integration strategy cannot succeed in the long term, as still further integration brings the need to deal with an unmanageable amount of information at the detailed level. Rather, integration which emphasizes vertical connections from the particular to the general within knowledge about an issue achieves the necessary economy when coupled with horizontal integration between issues at the general level. Some details are important enough that they must be dealt with on a detailed basis, but the few which are important are obscured when every issue is handled in the same way.

When human engineers use a tool, it must act as an extension of their own abilities. Use of the design tool must provide adequate feedback, allowing the engineer to verify the actions of the tool are in accordance with the needs of the current design decision. The simplest form of feedback is that the results of the tool's action may be verified independently. For instance, the results of a complex numerical model may be checked by using a less detailed (but more economical) analytic expression or from experimental data. Tools which produce evidence which is not easy to check are of dubious value; experienced engineers usually devise means to check the tools they use.

Since the tool is built to serve the engineer, it cannot be allowed to dominate the user. The human engineer is the only irreplaceable part of a design system; no tool or tool collection should be built with the intent that the engineer's needs can be bent to meet the limits of the tool. Human initiative cannot be supplanted by the limits of other systems.

One maxim can be used to summarize the proper attitude for a design tool developer: The developer can never anticipate how a tool will be used, and must never dictate how it should be used. Rather, the developer's assignment is to anticipate that many applications cannot be planned or predicted, and to build the tool with the properties required to assure it will continue to serve engineering needs through a wide variety of applications.

9.4. Inexpensive Tricks

In recent times there has been a tendency among some individuals to dismiss simplified approaches, such as those advocated in this thesis, as "cheap tricks". The author has heard these comments first-hand. The author has two responses to this comment. First, it is necessary to make a distinction between cheap tricks and inexpensive tricks. Not all inexpensive things are cheap; not all cheap things are inexpensive. Second, dismissing inexpensive tricks simply because they are inexpensive is a judgement, based entirely on fashion and aesthetics, which forces commitment to the use of only expensive tricks. Since managing the allocation of scarce resources is the main activity of engineering, it does not seem sensible to deliberately choose reasoning approaches which involve greater consumption of engineering time, one of the scarce resources to be managed.

Inexpensive tricks are important to engineers because economical problem-solving is vital to successful design. There are many design situations which can be solved through the use of simple, reasonably valid evaluations, avoiding the expense of a more complicated calculation or experiment. Known inaccuracies may be acceptable in the context of the immediate decision to be made. Inexpensive tricks are required for the engineer to be able to check the results obtained from tools and methods which deal with the problem on a more complicated scale. Inexpensive tricks allow the engineer to carry out some mental context-setting in order to more effectively interpret and apply the results obtained from more complex tools.

Inexpensive tricks are also important to the organizations which employ engineers. It is necessary for organizations losing their competitive edge to regain a foothold in the international marketplace. This cannot be accomplished by manufacturing improvements alone; in the author's experience a better product will outsell a cheaper product. Therefore better engineering is the primary aspect of winning back the ability to compete. Inexpensive tricks support better engineering by allowing the engineer to make effective decisions and to make design decisions in a minimum-commitment mode.

Frequently the apparent simplicity of an inexpensive trick in its final form conceals the complex and difficult process of developing that trick. This observation explains, in part, the confusion between cheap tricks and inexpensive tricks. A feature of an inexpensive trick is that much of the foundational, preparatory thinking has been done in advance. This, unfortunately, helps to preserve the illusion that the reasoning required is simply to blindly follow the "cookbook" formula. Inexpensive tricks are considerably more useful when the user understands the background; the trick relieves the user from resorting to first-principles reasoning with every problem. Judgement about the validity of an inexpensive trick is necessary every time it is used.

The word "cheap" is normally a pejorative, carrying with it the implication of poor quality or value. The author has seen offered as "design tools" a fair number of techniques and tools which, although expensive to obtain or to operate, were of low value in the context of most design situations. Not all cheap things are inexpensive.

Conversely, not all inexpensive things are cheap. One might assume a simple reasoning method has little value; frequently the opposite relationship holds and simplicity is a valuable feature. In this thesis a number of illustrations have been provided of situations where economical, simplified reasoning methods are of value during design. Which methods, which level of simplification, will be appropriate for a given design situation cannot be judged in advance by any standard, and an aesthetic standard of rigor is probably the worst standard which might be applied. The value of a particular method of solving a problem can only be evaluated in the context of an engineer trying to use it to solve a problem.

One must exercise caution, of course, to avoid inundating engineers with inexpensive techniques, just as they have been oversupplied with detailed and complex methods in the recent past. To assure the value of a technique is recognizable, its producer must set a context for its use: the limiting assumptions which apply, the validation which has been accomplished, and the problems to which it has been successfully applied.

In this thesis, one major theme has predominated: economy of problem solving by engineering designers. Three issues associated with economy have been examined: dealing with details in the context of a broader understanding of the problem and solution; providing reasoning and problem-solving techniques and tools which can be adapted to suit the needs of engineers making design decisions at a variety of levels of detail; and providing the means to make decisions which have a favorable effect on the overall quality of the design. All of these ideas are important to the broad goal of improving the productivity of engineering designers by making their decisions more effective. All of them require further development of inexpensive tricks.

9.5. Conclusion

A design cannot be moved forward to realization as a product without engineers taking the risk of using inexact solution methods, rules of thumb, even guessing. There are many methods of approximate reasoning which engineers use routinely. In this research, a small selection of reasoning methods for design have been explored. From this exploration, conclusions and recommendations may be drawn. These conclusions are answers to the research questions posed in the first chapter.

The flexibility of the engineering knowledge base can be improved by acknowledging the importance of reasoning and problem-solving methods. Methods in engineering are worthy of a research emphasis equal to that placed on developing new information. Simply developing new information is no longer enough: research results must be expressed in ways conducive to use as the basis for designs which include the new technology. Specific issues which much be addressed to improve the flexibility of the engineering knowledge base are:

- making knowledge more accessible by developing a broader, more generalized view of engineering knowledge,
- recognizing the value of approximate reasoning methods and making an explicit effort to develop estimating methods and rules of thumb for new information as it is generated, and
- generalizing detailed information in an effort to develop more concise forms of knowledge which can serve as a gateway to areas previously unfamiliar to the engineer

The coherence of the engineering knowledge base can be improved by recognizing the need for a context of prior knowledge to help the engineer understand and apply new knowledge in design. Specific issues which must be addressed to enhance the coherence of the engineering knowledge base are:

- making knowledge more accessible by developing the links between disciplines of research which are generally treated as discrete areas,
- focusing on "fitness for purpose" during research, recognizing that engineering research in intended to be applied during design shortly after the research is finished, and
- generalizing detailed information in an effort to have a broad base of concise knowledge which can be the foundation of the context-setting required to introduce new knowledge

The predictive value of engineering knowledge can be improved by recognizing that "inverse problems" are the standard mode of thought during design. There is a long-standing prejudice towards analysis of existing devices and systems. Most of the information learned during a baccalaureate program is suitable for feedback about a complete design, illustrated by the types of problems posed by the authors of most textbooks for undergraduate courses. This prejudice must be reversed. Knowledge presented to engineers must emphasize feed-forward, predictive methods of reasoning. Changing textbooks and the teaching methods using the textbooks takes time. Making the conversion to a design focus in every course offered will be worth the effort, because every student will develop a clearer understanding of design, and the risk-intensive methods necessary to accomplish design.

The problem with a feed-back, reactive design approach is that the first proposed solution is necessarily a dummy, an initial point from which to start iterating or searching. The iterations are reactions to the previous analysis. Good judgement based on experience is necessary to establish a first guess near the final solution. A feed-forward, pro-active design approach can eliminate some iteration altogether. When iteration is unavoidable, even the first point will be better selected. Each decision the engineer makes is based on a prediction of the future. Therefore predictive value is a very important property of design information.

The detailed knowledge currently available can be generalized to support earlier, more global decisions. In many cases, the effort required to generalize knowledge is not large. For example, the generalization of manufacturability information discussed in the fourth chapter followed a simple and straightforward reasoning path. The value of the exercise in Chapter 4 is more in the demonstration of a goal and a means to that goal, and much less so in the generalized knowledge obtained.

Economy of decision-making can be improved by providing engineers with the knowledge and tools necessary to make advances in both efficiency and effectiveness during design thinking. Efficiency is the main focus of the computer aids to engineering currently being sold; a majority of advertising claims deal with solving problems faster. It is not certain that engineering productivity can be greatly improved by relying solely on making faster decisions.

In some cases, problems are not being solved faster. Instead, problems are being solved in the same amount of time after considering more evidence than before. By examining evidence which incorporates a more complete picture of he situation, decisions can be made more effective. That is, decisions can be made with a greater likelihood of success. However, the tendency has been to provide engineers with increasing volumes of evidence without discriminating between high and low value. Simply accumulating more evidence before making a decision will reduce the efficiency with which decisions are made.

Two methods can improve effectiveness by allowing engineers to rely on more evidence to support their decisions: evaluating the sensitivity of the evidence, and improved management and manipulation of the evidence. Some evidence will have a substantial effect on a decision, other information will have only a mild influence. When making a decision, the engineer will give more weight to the evidence which is more important. A system or tool helping the engineer decide how to weight evidence would be very useful. The weight given to evidence is similar to, but not the same as, a formal sensitivity analysis. In both cases the goal is to determine the impact of a parameter's variability. But for evidence, the process is considerably more qualitative in flavor. Evidence may take the form of discrete points, making the differential associated with sensitivity analysis meaningless.

The tools currently on the market for computer aided engineering seem to be focused on generating more evidence but lack the ability to help the engineer discriminate between more and less important information. For this reason, judgement based on experience is a key component of effective design engineers. Development of general methods and specific tools to aid in managing, manipulating, and correlating evidence is needed.

Although the computer has clearly helped engineers, the computer has been mis-used as well, at the cost of reducing engineering productivity. Software tools which require large inputs of data and generate large outputs impede decision-making by requiring a slow search for significant evidence within the mass of results produced.

Engineers have ample support from tools built with the intent of improving productivity by increasing the pace at which decisions can be made. Most CAE tools are intended to be used to generate evidence. There is a strong need for tools which help the engineer to organize, sift, and correlate evidence in support of decision-making. As a corollary, all tools must deliver the evidence they generate in a form suitable for use as one of many inputs into each decision.

The computer as a basis for design tools is only beginning to be understood. The most important goal for developers of CAE is to have the computer support the engineer making decisions. The computer should not (and, this author believes, can not) make design decisions. The behavior of the computer and the software should not dominate the user's cognitive capacity. Tools to support engineering design must act as an extension of the engineer's intent and capabilities.

The clearest path towards improving engineering productivity is to improve the quality of the designs being produced without sacrificing (and hopefully improving) the efficiency with which decisions are made. In this research, some experiments were conducted dealing with new ways to think about design problem-solving. The most important conclusion made is simply to note that design engineers, every day, make decisions based on partial information to advance the design towards completion.

Engineering research has little value if its results cannot be used to design a better product. In the past, a long period of settling and "reduction to practice" has been permitted as the transition between research and application. With increasing emphasis on shorter product development cycles, and the need to take advantage of every advance in technology, such a casual link between research and design can no longer be tolerated. There will be an increasing pressure on designers to use new technology, and there should be an equivalent pressure on researchers to make new technology useful.

In addition to the need to better focus the products of research in the future, there are many years of research already complete which need an equivalent focusing treatment. Most research is best understood by the researchers who performed the work. Engineering researchers need to review their past work and develop the information generated into forms which make it more useful during design. Peer-reviewed, academic journal articles receive emphasis among university scholars at present. It may be appropriate to give equal emphasis to articles appearing in trade publications, for instance *Machine Design*, which explain the design value of research results.

There is a tradition of relying on evolution to introduce new technologies. Research is completed, results published, and the sponsors of the research may achieve some tangible benefit in a product. Later, the understanding seeps slowly out to the general population of engineers. Typically, it takes years before a research result is commonly understood and applied in design.

Random processes are no longer an adequate means of moving information from researchers to design engineers. Researchers must shoulder the burden of making their results more meaningful. However, researchers cannot make a substantial move toward improving design emphasis unless the standards by which their performance is evaluated are also changed. Sponsors and managers of researchers must acknowledge the importance of improved communications from research results to design applications, and make changes which encourage the necessary changes.

The single most important change necessary is to develop design tools and conduct engineering research with the intent that the results will provide support to engineering decision-making. Design tool developers must focus on what techniques and methods fit the ways engineers work. Researchers must present their results in ways that allow the new information to fit well in with the existing base of knowledge already available. Researchers must take the time to better understand how engineering decisions are made.

References

Abeyama, S., Kimura, A., and Nakamura, S., "The Influence of Heat Treatment and Cold Forging on Machinability of Low-Alloyed Steels", in Thompson, R.W. (Editor), *The Machinability of Engineering Materials*, American Society for Metals, 1983.

Agarwal, B.D., and Broutman, L.J., *Analysis and Performance of Fiber Composites*, Wiley-Interscience, 1980.

Aries Technology, "Desktop Systems for Mechanical Engineers", product literature, 1990.

Akin, O., "How do Architects Design?", in Latombe, J.-C. (Editor), *Artificial Intelligence and Pattern Recognition in Computer-Aided Design*, North-Holland, 1978

Alcoa Structural Handbook, Aluminum Company of America, 1960.

Alexander, C., *Notes on the Synthesis of Form*, Harvard University Press, 1968.

Alexander, E.R., "After Rationality, What?", *Journal of the American Planning Association*, Volume 50, number 1, 1984.

Alexander, E.R., "Design in the Decision-Making Process", *Policy Sciences*, Volume 14, number 3, June 1982.

Alic, J.A., "The Rise and Fall of R&D", *Mechanical Engineering*, Volume 110, number 4, April 1988.

Alger, J.R.M., and Hays, C.V., *Creative Synthesis in Design*, Prentice-Hall, 1964.

Algor Inc., Iconnex Parametric Modeling, product literature, 1990.

Altman, S.M., "How Long Can we Go On This Way?", in Spillers, W.R. (Editor), *Basic Questions of Design Theory*, North-Holland/American Elsevier, 1974.

Araki, T., Fukunaga, H., Sata, T., and Narutaki, N., "Some Results of Cooperative Research on the Effect of Heat-Treated Structure on the Machinability of a Low-Alloy Steel", in Tipnis, V.A. (Editor), *Influence of Metallurgy on Machinability*, American Society for Metals, 1975.

Argyris, J.H., St. Doltsinis, J., and Wustenberg, H., "Analysis of Thermoplastic Forming Processes - Natural Approach", in Pittman, J.F.T., Zienkiewicz, O.C., Wood, R.D., and Alexander, J.M., *Numerical Analysis of Forming Processes*, Wiley, 1984.

Ashley, S., "The Battle to Build Better Products", *Mechanical Engineering*, Volume 112, number 1, November 1990.

Asimow, M., *Introduction to Design*, Prentice-Hall, 1962.

Baer, T., "Simulating the Factory", *Mechanical Engineering*, Volume 108, number 2, December 1986.

Bailey, R.L., *Disciplined Creativity for Engineers*, Ann Arbor Science/Butterworth, 1978.

Ballem, H., "Identification and Analysis of Specific Problem-Solving Strategies", in Shufelt, G., and Smart, J.R. (Editors), *The Agenda in Action*, Yearbook of the National Council of Teachers of Mathematics, 1983.

Barrett, A.J., "The Refinement of Data Resources to Assist Engineering Design", in *Engineering Research and Design - Bridging the Gap*, Institution of Mechanical Engineers (London), 1981. Paper C231/81, publication 1981-7.

Bazjanac, V., "Architectural Design Theory: Models of the Design Process", in Spillers, W.R. (Editor), *Basic Questions of Design Theory*, North-Holland/American Elsevier, 1974.

Beakley, G.C., and Leach, H.W., *Engineering: an Introduction to a Creative Profession*, Third Edition, Macmillan, 1977.

Bednarz, S.M., *Development of a Failure Database for Mechanical Reliability Improvement*, M.S. Thesis, Department of Mechanical and Industrial Engineering, University of Illinois at Urbana-Champaign, 1988.

Bednarz, S.M., and Marriott, D.L., "Efficient Analysis for FMEA", Annual Reliability and Maintainability Symposium, Los Angeles, January 1988, Institute of Electrical and Electronic Engineers, 1988.

Begeman, M.L., and Conkon, J., "The Right Tool for the Job", *Byte*, Volume 13, number 10, October 1988.

Begg, V., *Building Expert CAD Systems*, Kogan Page, 1984.

Berma, I., "The Art, Science, and Practice of Mechanical Design", *Mechanical Engineering*, Volume 108, number 8, August 1986.

Berry, D.T., "Stamping Out Forming Problems with FEA", *Mechanical Engineering*, Volume 110, number 7, July 1988.

Bhattacharyya, R., *Dynamics of Marine Vehicles*, Wiley, 1978.

Blackstone, C., "The Future of the Design Engineer in the U.K.", in Davies, B.J. (Editor), Seventh Annual Design Engineering Conference, The Design Council (London)/North-Holland, 1984.

Blair, D.C., and Maron, M.E., "An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System", *Communications of the A.C.M.*, Volume 28, number 3, March 1983.

Blauth, R.E., and Preston, E.J., "Computing the Payback for CAD", *Machine Design*, Volume 53, number 19, August 1981.

Block, H.P. and Geitner, F.K., *An Introduction to Machinery Reliability Assessment*, Van Nostrand Reinhold, 1990.

Blodgett, O.W., *Design of Weldments*, James F. Lincoln Arc Welding Foundation, 1963.

Bowell, J.C., "CAD/CAM - Successful Implementation", in Davies, B.J. (Editor), Seventh Annual Design Engineering Conference, The Design Council (London)/North-Holland, 1984.

Box, G.E.P., and Biggaard, S., "The Scientific Context of Quality Improvement", *Quality Progress*, Volume 20, number 6, 1987.

Braham, R., "Opting Out: Putting an End to Over-Design", *Mechanical Engineering*, Volume 110, number 2, February 1988.

Brome, R., and Brophy, J., "Teachers' Cognitive Activities", in Christiansen, B., Hanson, A.G., and Otte, M. (Editors), *Perspectives on Mathematics Education*, D. Reidel, 1986.

Brosseau, G., Davis, R.B., and Werner, T., "Observing Students at Work", in Christiansen, B., Hanson, A.G., and Otte, M. (Editors), *Perspectives on Mathematics Education*, D. Reidel, 1986.

Brown, D.C., and Breau, R., "Types of Constraints in Routine Design Problem-Solving", in Sriram, D., and Adey, R. (Editors), *Applications of Artificial Intelligence in Engineering Problems*, Springer-Verlag, 1986.

Brown, J.S., "The Low Road, the Middle Road, and the High Road", in Winston, P.H., and Prendergast, K.A., *The A.I. Business*, MIT Press, 1984.

Buchanan, B.G., "New Research on Expert Systems", in Hayes, J.E., Michie, D., and Pao, Y.-H. (Editors), *Machine Intelligence 10*, Halstead (Wiley), 1982.

Bulkley, P.Z., "The 1967 Summer Institute in Interdisciplinary Design", in Fourth Conference on Engineering Design: Engineering Design and Design Education, Hanover, New Hampshire, July 1967.

Butcher, J.C., and Ensor, P.S., "A Tire Maker Retreads its Corporate Structure", *Mechanical Engineering*, Volume 110, number 4, April, 1988.

Campbell, J.R., "Avoiding Fracture in Pressure Vessels", in First National Conference on Fracture, University of Witwatersrand, Johannesburg, 1979.

Carbonell, J.G., "Learning by Analogy: Formulating and Generalizing Plans from Past Experience", in Michalski, R.S., Carbonell, J.G., and Mitchell, T.M. (Editors), *Machine Learning: An Artificial Intelligence Approach*, Morgan Kaufman, 1983.

Chang, D., "PC's Offer More Options for Less Money", *Mechanical Engineering*, Volume 111, number 5, May, 1989.

Chorafas, D.N., *Engineering Productivity Through CAD/CAM*, Butterworth, 1987.

Christie, B. (Editor), *Human Factors of the User-System Interface*, North-Holland, 1985.

Christopherson, D.G., "Opening Address: Discovering Designers", in Jones, J.C., and Thornley, D.G., (Editors), *Conference on Design Methods*, Macmillan, 1963.

Cognition, Inc., "Mechanical Advantage: the Next Logical Step", product literature, 1987.

Collins, J.A., *Failure of Materials in Mechanical Design*, Wiley, 1981.

Comer, J.J., *An Investigation into the Skills Necessary for the Success of Newly Graduated Engineers in Industry*, Ph.D. thesis, Department of Mechanical and Industrial Engineering, University of Illinois at Urbana-Champaign, 1987.

Computervision, "Design View", product literature, 1991.

Comstock, J.P. (Editor), *Principles of Naval Architecture*, Society of Naval Architects and Marine Engineers, 1967.

Conklin, J., "Hypertext: An Introduction and Survey", *Computer*, Institute of Electrical and Electronic Engineers, September 1987.

Cook, N.H., "What is Machinability?", in Tipnis, V.A. (Editor), *Influence of Metallurgy on Machinability*, American Society for Metals, 1975.

Cooley, M.J.E., "Problems of Automation", in Lupton, T., (Editor), *Human Factors: Man, Machine, and New Technology*, IFS (Publications) and Springer-Verlag, 1986.

Coombs, M.J., "Artificial Intelligence and Cognitive Technology: Foundations and Prospects", in Hollnagel, E., Mancini, G., and Woods, D.D. (Editors), *Intelligent Decision Support in Process Environments*, Springer-Verlag, NATO ASI series, Volume F21, 1986.

Coombs, M.J., and Alty, J.L., "Expert Systems: An Alternative Paradigm", *International Journal of Man-Machine Studies*, Volume 20, 1984.

Crosby, P.B., *Quality is Free*, McGraw-Hill, 1979.

Cross, H., *Engineers and Ivory Towers*, McGraw-Hill, 1942.

Cross, N., "The Common Core of Design", in Fourth International Conference on the Education and Training of Engineering Designers, Institute of Mechanical Engineers (London), 1985.

Datsko, J., *Material Properties and Manufacturing Processes*, Wiley, 1966.

De Bono, E., "Beyond Critical Thinking", *Curriculum Review*, Volume 25, number 3, January/February 1986.

Decker, R.W., "Keynote Address: Statistical Process Control at General Motors", *Statistical Process Control*, Society of Automotive Engineers, publication SP-547, June 1983.

Deitz, D., "The Power of Parametrics", *Mechanical Engineering*, Volume 111, number 1, January 1989.

De Marco, T. *Structured Analysis and System Specification*, Yourdon/Prentice-Hall, 1979.

Dieter, G.E., *Engineering Design: a Materials and Processing Approach*, McGraw-Hill, 1983.

Dieter, G.E., *Mechanical Metallurgy*, Third Edition, McGraw-Hill, 1986.

Dillon, J.T., "Problem Finding and Solving", *Journal of Creative Behavior*, Volume 10, number 2, 1982.

Dimarogonas, A.D., *Computer Aided Machine Design*, Prentice-Hall, 1988.

Dixon, J.R., *Design Engineering*, McGraw-Hill, 1966.

Doblin, J., "Information and Design: the Essential Relation", in Whitney, P. (Editor), *Design in the Information Environment*, Knopf, 1986.

Donovan, J.C., "Creative Success: Twenty Years on the Back of an Envelope", *Journal of Creative Behavior*, Volume 19, number 2, 1985.

Dorfler, W., and McLowe, R.R., "Mathematics as a School Subject", in Christiansen, B., Howson, A.G., and Otte, M. (Editors), *Perspectives on Mathematics Education*, D. Reidel, 1986.

Downing, T., "Eight New Ways to Evaluate Automation", *Mechanical Engineering*, Volume 111, number 7, July 1989.

Dreyfus, H.L., and Dreyfus, S.E., *Mind Over Machine*, The Free Press, 1986.

Drozda, T.J., and Wick, C, (Editors), *Tool and Manufacturing Engineers Handbook*, Volume I: Machining, Fourth Edition, Society of Manufacturing Engineers, 1983.

Easterling, K., *Introduction to the Physical Metallurgy of Welding*, Butterworth, 1983.

Eastman, C.M., "The Representation of Design Problems and Maintenance of their Structure", in Latombe, J.-C. (Editor), *Artificial Intelligence and Pattern Recognition in Computer-Aided Design*, North-Holland, 1978

Eddy, P., Potter, E., and Page, B., *Destination Disaster*, Hart-Davis, MacGibbon, 1976.

Eder, W.E., "Engineering Design - Practice and Theory", Design Engineering Technical Conference, American Society of Mechanical Engineers, Paper 86-DET-113, 1986.

Esherick, J., "Problems of the Design of a Design System", in Jones, J.C. and Thornley, D.G. (Editors), *Conference on Design Methods*, MacMillan, 1963.

Evans, B., "Simultaneous Engineering", *Mechanical Engineering*, Volume 110, number 2, February 1988.

Evanson, S.E., "How to Talk to an Expert", *AI Expert*, Volume 3, number 2, February 1988.

Fabrycky, W.J., "Designing for the Life Cycle", *Mechanical Engineering*, Volume 109, number 1, January 1987.

"Failure Analysis and Prevention", *Metals Handbook*, Ninth Edition, Volume 11, American Society for Metals, 1986.

Faupel, J.H., and Fisher, F.E., *Engineering Design*, Second Edition, Wiley, 1981.

Fenwell, F.M., "Focusing on Problem Solving in the Primary Grades", in Shufelt, G., and Smart, J.R. (Editors), *The Agenda in Action*, Yearbook of the National Council of Teachers of Mathematics, 1983.

Fiderio, J., "A Grand Vision", *Byte*, Volume 13, number 10, October 1988.

Finegold, A., "The Engineer's Apprentice", in Winston, P.H., and Prendergast, K.A., *The A.I. Business*, MIT Press, 1984.

Fischhoff, B., Slovic, P., and Lichtenstein, S., "Fault Trees: Sensitivity of Estimated Failure Probabilities to Problem Representation", *Journal of Experimental Psychology: Human Perception and Performance*, Volume 4, 1978.

Foar, A.R., "If Engineering Software Were as Popular as Arcade Games ...", in Davies, B.J. (Editor), Seventh Annual Design Engineering Conference, The Design Council (London)/North-Holland, 1984.

Fouet, J.-M., "Artificial Intelligence in Structural Analysis", in Niku-Lavi, A. (Editor), "Finite, Boundary Element, and Expert Systems in Structural Analysis", *Structural Analysis Systems*, Volume 5, Pergamon, 1986.

Frankel, A., "What is the Design Process", in Furman, T.T. (Editor), *The Use of Computers in Engineering Design*, Van Nostrand Reinhold, 1970.

Freeman, G., "2-D Draughting - the New Approach", in Davies, B.J. (Editor), Seventh Annual Design Engineering Conference, The Design Council (London)/North-Holland, 1984.

French, M.J., *Conceptual Design for Engineers*, Second Edition, Springer-Verlag/The Design Council (London), 1985.

Frisse, M., "From Text to Hypertext", *Byte*, Volume 13, number 10, October 1988.

Fuchs, H.O., "Teaching Design by Precept, Practice, and Post-Mortem", Fourth National Conference on Engineering Design, Dartmouth College, 1967.

Gaertner, G.H., Nieva, V.F., and Argoff, N.J., "Managing to Preserve the Institutional Memory", *Mechanical Engineering*, Volume 108, number 6, June 1986.

Gall, J., *Systemantics: How Systems Work and Especially How they Fail*, Quadrangle, 1977.

Gellatly, A., "Cognition and Psychology", in Gellatly, A. (Editor), *The Skilful Mind*, Open University Press, 1986.

Gero, J.S., "Ethics in Computer-Aided Design: a Polemic", in Spillers, W.R. (Editor), *Basic Questions of Design Theory*, North-Holland/American Elsevier, 1974.

Gero, J.S. (Editor), *Knowledge Engineering in Computer-AIded Design*, North-Holland, 1985.

Gero, J.S., and Balachandran, M., "Knowledge and Design Decision Processes", in Sriram, D., and Adey, R. (Editors), *Applications of Artificial Intelligence in Engineering Problems*, Springer-Verlag, 1986.

Gero, J.S., Radford, A.D., Coyne, R., and Alines, V.T., "Knowledge-Based Computer-Aided Architectural Design", in Gero, J.S. (Editor), *Knowledge Engineering in Computer-AIded Design*, North-Holland, 1985.

Gettys, C.F., Manning, C.A., Mehle, T., and Fisher, S., *Hypothesis Generation: A Final Report of Three Years of Research*, Report TR 15-10--80, Decision Process Laboratory, Department of Psychology, University of Oklahoma, 1980.

Glegg, G.L., *Design of Design*, Cambridge University Press, 1969.

Goldstein, G., "Integrating Product and Process Design", *Mechanical Engineering*, Volume 111, number 4, April 1989.

Gordon, W.J.J., *Synectics: the Development of Creative Capacity*, Harper and Row, 1961.

Gosling, W., *Design of Engineering Systems*, Heywood, 1962.

Graville, B.A., *The Principles of Cold Cracking Control in Welds*, Dominion Bridge Company (Montreal), 1975.

Gregory, S.A., and Turner, B.T., "The Scope of Engineering, Including Technical and Social Models", in Gregory, S.A. (Editor), *Creativity and Innovation in Engineering*, Butterworth, 1972.

Grosskreutz, J.C., and Mughrabi, H., "Description of the Work-Hardened Structure at Low Temperature in Cyclic Deformation", in Argon, A.S. (Editor), *Constitutive Equations in Plasticity*, M.I.T. Press, 1975.

Hall, A.D., *A Methodology for Systems Engineering*, Van Nostrand, 1962.

Halpern, S., *The Assurance Sciences: An Introduction to Quality Control and Reliability*, Prentice-Hall, 1978.

Hammond, N., and Barnard, P., "Dialogue Design: Characteristics of User Knowledge", in Monk, A. (Editor), *Fundamentals of Human-Computer Interaction*, Academic press, 1984.

Handrock, J.L., *Development of Material Failure Logic Models for Component Design, Manufacture, and Evaluation*, M.S. Thesis, Department of Mechanical and Industrial Engineering, University of Illinois at Urbana-Champaign, 1984.

Hankins, G., "The Paradox of Expertise", *Engineering Education*, Volume 77, number 5, February 1987.

Harris, L.R., "Natural Language Front Ends", in Winston, P.H., and Prendergast, K.A., *The A.I. Business*, MIT Press, 1984.

Hatvany, J., "The Missing Tools of CAD for Mechanical Design", in Gero, J.S. (Editor), *Knowledge Engineering in Computer-AIded Design*, North-Holland, 1985.

Hayes-Roth, F., Waterman, D., and Lenat, D. (Editors), *Building Expert Systems*, Addison-Wesley, 1983.

Heine, S., and Rohr, G., "Human-Computer Interaction: A Problem of Cognitive Ergonomics", in Willumeit, H.-P. (Editor), *Human Decision-Making and Manual Control*, Elsevier, 1986.

Henley, E.J., and Kumamoto, H., *Reliability Engineering and Risk Assessment*, Prentice-Hall, 1981.

Heuschkel, J., "Steel Properties Related to Welded Performance", *Welding Journal*, Volume 38, number 3, March 1949.

Hillman, D.J., "Artificial Intelligence", *Human Factors*, Volume 27, number 1, 1985

Hofstadter, D.R., *Godel, Escher, Bach: an Eternal Golden Braid*, Basic Books, 1979.

Holland, J., "The Historian, the Critic, and the Designer", in *Design History: Fad or Function?*, The Design Council (London), 1978.

Horn, D., "Expert Systems Emerge from Their Shells", *Mechanical Engineering*, Volume 111, number 4, April 1989.

Hosford, W.F., and Caddell, R.N., *Metal Forming: Mechanics and Metallurgy*, Prentice-Hall, 1983.

House, P.A., Wallace, M.L., and Johnson, M.A., "Problem-Solving as a Focus: How? When? Whose Responsibility?", in Shufelt, G., and Smart, J.R. (Editors), *The Agenda in Action*, Yearbook of the National Council of Teachers of Mathematics, 1983.

Hubka, V. (W.E. Eder, Translator), *Principles of Engineering Design*, Butterworth Scientific, 1982.

Hunter, I.M.L., "Exceptional Memory Skills", in Gellatly, A. (Editor), *The Skilful Mind*, Open University Press, 1986.

Hunter, T.A., "After the Fact ...", *Mechanical Engineering*, Volume 111, number 3, March 1989.

Hutchings, F.R., and Unterweiser, P.M. (Editors), *Failure Analysis: The British Engine Technical Reports*, American Society for Metals, 1981.

Jones, J.C., "A Method of Systematic Design", in Jones, J.C., and Thornley, D.G., (Editors), *Conference on Design Methods*, Macmillan, 1963.

Joyce, J.D., and Uthurusamy, R., "Promises and Pitfalls of Knowledge Systems: an Overview", SAE International Congress and Exposition, Society of Automotive Engineers, Paper 860333, 1986.

Juvinall, R.C., *Fundamentals of Machine Component Design*, Wiley, 1983.

Kahneman, D., Slovic, P., and Tversky, A., *Judgement Under Uncertainty: Heuristics and Biases*, Cambridge University Press, 1982.

Kahneman, D., and Tversky, A., "Subjective Probability: A Judgement of Representativeness", *Cognitive Psychology*, Volume 3, 1972.

Kantowksi, M.G., "Problem Solving", in Fennema, E. (Editor), *Mathematics Education Research: Implications for the 80's*, Association for Supervision and Curriculum Development/National Council of Teachers of Mathematics, 1981.

Kidd, A., "Human Factors Problems in the Design and Use of Expert Systems", in Monk, A. (Editor), *Fundamentals of Human-Computer Interaction*, Academic press, 1984.

Knight, F.L., "Computer-Aided Design Software for Mechanical Engineering Design", in 1983 International Computers in Engineering Conference and Exhibit, Computers in Engineering Division, American Society of Mechanical Engineers, August 7-11, 1983, Chicago.

Knoll, F., "Checking Techniques", in Nowack, A.S. (Editor), *Modeling Human Error in Structural Design and Construction*, American Society of Civil Engineers, 1986.

Koen, B.V., *Definition of the Engineering Method*, American Society for Engineering Education, 1985.

Kuzela, L., "'CAD/CAM Reduce Creativity'", *Industry Week*, Volume 238, number 9, May 1, 1989.

Laberge, D., and Samuels, S.J., "Toward a Theory of Automatic Information Processing in Reading", *Cognitive Psychology*, Volume 6, number 2, 1974.

Latombe, J.-C. (Editor), *Artificial Intelligence and Pattern Recognition in Computer-Aided Design*, North-Holland, 1978

Leontief, W., "The Ins and Outs of Input/Output Analysis", *Mechanical Engineering*, Volume 109, number 1, January 1987.

LeShan, L., "How Can You Tell a Physicist from a Mystic?", *Intellectual Digest*, February 1972.

Liebowitz, H. (Editor), *Fracture: an Advanced Treatise*, Seven Volumes, Academic Press, 1969

Lind, M.C., "Control of Human Error in Structures", in Nowack, A.S. (Editor), *Modeling Human Errors in Structural Design and Construction*, American Society of Civil Engineers, 1986.

Lindberg, R.A., *Processes and Materials of Manufacture*, Allyn and Bacon, 1964.

Lockhead, G.R., "Know, Then Decide", in Wallsten, T.S. (Editor), *Cognitive Processes in Choice and Decision Behavior*, Erlbaum, 1980.

Love, S.F., *Planning and Creating Successful Engineered Solutions*, Van Nostrand Reinhold, 1980.

Lue, W.-R., *A Computer Implementation of the Material Failure Logic Model Technique*, Ph.D. Thesis, Department of Mechanical and Industrial Engineering, University of Illinois, 1985.

MCAE Technologies, Inc., "CEDAR: Personal Software for 'What-If' Engineering", Product Literature, 1990.

Mach, E., *The Science of Mechanics: A Critical and Historical Account of its Development*, Sixth American Edition, T.J. McCormack (Translator), Open Court, 1960.

Machining Data Handbook, Second Edition, Machinability Data Center (Metcut Research Associates), 1972.

Maher, M.L., and Fenves, S.J., "HI-RISE: An Expert System for the Preliminary Structural Design of High Rise Buildings", in Gero, J.S. (Editor), *Knowledge Engineering in Computer-AIded Design*, North-Holland, 1985.

Mandelbaum, S.J., "A Complete General Theory of Planning is Impossible", *Policy Sciences*, Volume 11, 1979.

McKim, R.H., "Visual Thinking and the Design Process", Fourth National Conference on Engineering Design, Dartmouth College, 1967.

Marriott, D.L., and Miller, N.R., "Material Failure Logic Models - A Procedure for Systematic Identification of Material Failure Modes in Mechanical Components", *Transactions*, Volume 104, American Society of Mechanical Engineers, 1982.

Marriott, D.L., "Efficient Representation of Fracture Mechanics Knowledge for Design Applications", Materials and Engineering Design Conference 1988, Institute of Metals (London), 1988.

Marshall, R.D., Pfrang, E.O., Leyendecker, E.V., Woodward, K.A., Reed, R.P., Kasen, M.B., Shives, T.R., *Investigation of the Kansas City Hyatt Regency Walkways Collapse*, U.S. Department of Commerce, National Bureau of Standards, NBSIR 82-2465, February 1982.

McCrory, R.J., "Elements of Realism in Mechanical Reliability", American Society of Mechanical Engineers, Design Engineering Conference, 1965.

McQuaid, H.W., "The Metallurgical Phase of Fatigue Failures", Proceedings of the Society for Experimental Stress Analysis, Volume, number 3, 1946.

Medland, A.J., and Jones, A., "An Approach to Design Based Upon Functional Logic", in Davies, B.J. (Editor), Seventh Annual Design Engineering Conference, The Design Council (London)/North-Holland, 1984.

Meredith, D.D., Wong, K.W., Woodhead, R.W., and Wortman, R.H., *Design and Planning of Engineering Systems*, Second Edition, Prentice-Hall, 1985.

Michalski, R.S., and Stepp, R., "Revealing Conceptual Structure in Data by Inductive Inference", in Hayes, J.E., Michie, D., and Pao, Y.-H. (Editors), *Machine Intelligence 10*, Halstead (Wiley), 1982.

Michie, D., "A Theory of Advice", in Elcock, E.W. and Michie, D. (Editors), *Machine Intelligence 8*, Ellis Horwood, 1977.

Middendorf, W.H., *Design of Devices and Systems*, Marcel Dekker, 1986.

Miller, G.A., "The Magical Number Seven Plus or Minus Two: Some Limitations on Our Capacity to Process Information", *Psychological Review*, Volume 63, number 2, 1956.

Moray, N., "Modelling Cognitive Activities: Human Limitation in Relation to Computer Aids", in Hollnagel, E., Mancini, G., and Woods, D.D. (Editors), *Intelligent Decision Support in Process Environments*, Springer-Verlag, NATO ASI series, Volume F21, 1986.

Morris, K.W., "The Morphological Approach to Engineering Design", in Jones, J.C., and Thornley, D.G., (Editors), *Conference on Design Methods*, Macmillan, 1963.

Musiol, C., Draper, J., Sykes, N., and Morton, K., "Advances in Computer-aided Design Against Fatigue", in *Engineering Research and Design - Bridging the Gap*, Institution of Mechanical Engineers (London), 1981. Paper C234/81, publication 1981-7.

Nadler, G., "An Investigation of Design Methodology", *Management Science*, Volume 13, number 10, June 1967.

Nadler, G., *The Planning and Design Approach*, Wiley, 1981.

Nadler, G., "Systems Methodology and Design", *Mechanical Engineering*, Volume 108, number 9, September 1986.

Samuel Neaman Institute, "Engineering Education 2001", *Engineering Education*, Volume 78, number 2, November 1987.

Nevill, G.E., Jr., and Crowe, R.A., "Computer Augmented Conceptual Design", in Spiller, W.R. (Editor), *Basic Questions of Design Theory*, North-Holland/American Elsevier, 1974.

Newell, A., "Heuristic Programming: Ill-Structured Problems", in Aronofsky, J.S. (Editor), *Progress in Operations Research*, Volume III, Wiley, 1969.

Nitta, Y., Nukada, K., Mori, K., and Ohkawa, M., "Information Retrieval for Design Review Support", Annual Reliability and Maintainability Symposium, Institute of Electrical and Electronic Engineers, 1980.

Obrzut, J.J., "Materials Selection: How the Computer Helps Reduce Failures", *Iron Age*, April 27, 1981.

Olsen, S.E., and Andersen, H.B., "On the Use of Metacognition in Problem-Solving and in the Interpretation of Protocols", in Willumeit, H.-P. (Editor), *Human Decision-Making and Manual Control*, Elsevier, 1986.

Otte, M., "What is a Text?", in Christiansen, B., Hanson, A.G., and Otte, M. (Editors), *Perspectives on Mathematics Education*, D. Reidel, 1986.

Owen, D., "Answers First, then Questions", in Norman, D.A., and Draper, S.W. (Editors), *User-Centered System Design*, Erlbaum, 1986.

Pask, G., "The Conception of a Shape and the Evolution of a Design", in Jones, J.C., and Thornley, D.G., (Editors), *Conference on Design Methods*, Macmillan, 1963.

Patterson, C., "On the Efficient use of Numerical Model Data", in *Engineering Research and Design - Bridging the Gap*, Institution of Mechanical Engineers (London), 1981. Paper C232/81, publication 1981-7.

Peters, J., "AMT, a Problem of Communication Between Human Beings", *International Journal of Advanced Manufacturing Technology*, Volume 2, number 2, May 1987.

Peters, T.J., and Waterman, R.H., *In Search of Excellence*, Warner, 1984.

Petroski, H., *To Engineer is Human*, St. Martin's, 1985.

Polya, G., *How to Solve It*, Second Edition, Princeton University Press, 1957.

Price, H.E., "The Allocation of Functions in Systems", *Human Factors*, Volume 27, number 1, 1985.

Purcell, P.A., Maller, G.L., and Gomain, P.G.R., "A Strategy for Design Research", in Spillers, W.R. (Editor), *Basic Questions of Design Theory*, North-Holland/American Elsevier, 1974.

Richards, T.H., "Producing Stress and Vibration Analysis Data for Engineering Design", in *Engineering Research and Design - Bridging the Gap*, Institution of Mechanical Engineers (London), 1981. Paper C233/81, publication 1981-7.

Rittel, H.W.J., and Webber, M.M., "Dilemmas in a General theory of Planning", *Policy Sciences*, Volume 4, number 2, June 1973.

Rodenberger, C.A., Herndon, C.F., Majors, S.O., and Rogers, W.A., "The Average $100,000 Design Engineer", in Proceedings of the 1983 International Computers in Engineering Conference and Exhibit, Computers in Engineering Division, American Society of Mechanical Engineers, August 7-11, 1983, Chicago.

("The Rogers Commission") Presidential Commission on the Space Shuttle Challenger Accident, *Report to the President*, Washington, D.C., 1986.

Rogerson, J.H., "Defects in Welds - Their Prevention and Their Significance", in Burgess, N.T., (Editor), *Quality Assurance of Welded Construction*, Applied Science Publishers, 1983.

Roszak, T., *The Cult of Information*, Pantheon, 1986.

Rouse, W.B., and Boff, K.R. (Editors), *System Design: Behavioral Perspectives on Designers, Tools, and Organizations*, North-Holland, 1987.

Ryder, F.L., *Creative Engineering Analysis*, Prentice-Hall, 1961.

Saltire Software, "Analytix," Product Literature, 1990.

Sandler, B.-Z., *Creative Machine Design*, Paragon, 1985.

Schank, R.C., "Intelligent Advisory Systems", in Winston, P.H., and Prendergast, K.A. (Editors), *The AI Business*, MIT Press, 1984.

Schaeffler, A.L., "A Constitution Diagram for Stainless Steel Weld Metal", *Metals Progress*, Volume 58, number 6, November 1949.

Schraft, R.D., "The Importance of Assembly-Oriented Product Design - Examples and Possibilities to Realize", in Davies, B.J. (Editor), Seventh Annual Design Engineering Conference, The Design Council (London)/North-Holland, 1984.

Seireg, A.A., "Establishing the Market Value of Innovation", *Mechanical Engineering*, Volume 109, number 1, January 1987.

Shigley, J.E., and Mitchell, L.D., *Mechanical Engineering Design*, Fourth Edition, McGraw-Hill, 1983.

Siebert, O.W., "Classic Blunders in Corrosion Protection", *Materials Performance*, April 1978.

Simon, H.A., "The Structure of Ill-Structured Problems", *Artificial Intelligence*, Volume 4, 1973.

St. Denis, M., "Recollection of Some Personal Experiences in Ship Design and Research - Possibly Amusing or Embarrassing", *Marine Technology*, Volume 26, number 2, Society of Naval Architects and Marine Engineers, April 1989.

Starr, M.K., *Product Design and Decision Theory*, Prentice-Hall, 1963.

Stromboni, J.P., "How Can we Used Models in Man-Machine Systems Studies?", in Willumeit, H.-P. (Editor), *Human Decision-Making and Manual Control*, Elsevier, 1986.

Sussman, G.J., "SLICES: at the Boundary Between Analysis and Synthesis", in Latombe, J.-C. (Editor), *Artificial Intelligence and Pattern Recognition in Computer-Aided Design*, North-Holland, 1978

Swift, K.G., Matthews, A., and Syan, C.S., "The Application of IKBS in Design for Assembly and Surface Treatment Selection", in Sriram, D., and Adey, R. (Editors), *Applications of Artificial Intelligence in Engineering Problems*, Springer-Verlag, 1986.

System Dynamics Research Corporation, I-DEAS, product literature, 1990.

Szalapaj, P.J., and Bijl, A., "Knowing Where to Draw the Line", in Gero, J.S. (Editor), *Knowledge Engineering in Computer-AIded Design*, North-Holland, 1985.

Taguchi, G., and Wu, Y.-I., *Introduction to Off-Line Quality Control*, Central Japan Quality Control Association, 1979.

Taylor, W.A., "Development of a Knowledge Based System for Process Planning in Arc Welding", in Sriram, D., and Adey, R. (Editors), *Applications of Artificial Intelligence in Engineering Problems*, Springer-Verlag, 1986.

Teague, W.D., *Design This Day*, Harcourt, Brace, 1940.

Terry, G.J., "Methodology - It's Place in Computer-Aided Design", in Furman, T.T. (Editor), *The Use of Computers in Engineering Design*, Van Nostrand Reinhold, 1970.

Thimbleby, E., "User Interface Design: Generative User Engineering Principles", in Monk, A. (Editor), *Fundamentals of Human-Computer Interaction*, Academic Press, 1984.

Thompson, J., "New BEng and MEng Courses", in Fourth International Conference on the Education and Training of Engineering Designers, Institute of Mechanical Engineers (London), 1985.

Thomson, N., "Human Memory: Different Stores with Different Characteristics", in Monk, A. (Editor), *Fundamentals of Human-Computer Interaction*, Academic Press, 1984.

Timoshenko, S.P., and Goodier, J.N., *Theory of Elasticity*, Third Edition, McGraw-Hill, 1970.

The Timken Company, *Evaluating the Forgeability of Steel*, Fourth Edition, 1974.

Tomiyama, T., and Yoshikawa, H., "Requirements and Principles for Intelligent CAD Systems", in Gero, J.S. (Editor), *Knowledge Engineering in Computer-AIded Design*, North-Holland, 1985.

Tomiyama, T., and Yoshikawa, H., "Knowledge Engineering and CAD", in Yoshikawa, H., (Editor), *Design and Synthesis*, Proceedings of an International Symposium on Design and Synthesis, Tokyo, July 1984, Elsevier, 1985.

Tomiyama, T., and Yoshikawa, H., "Extended General Design Theory", in Yoshikawa, H., and Warman, E.A. (Editors), *Design Theory for CAD*, Elsevier/North-Holland, 1987.

Tucker, L.E., Landgraf, R.W., and Brose, W.R., "Proposed Technical Report on Fatigue Properties for the SAE Handbook", Society of Automotive Engineers, Paper 740279, 1974.

Turkstra, C.J., "The Case for Case Studies in Engineering Education", in Nowack, A.S. (Editor), *Modeling Human Error in Structural design and Construction*, American Society of Civil Engineers, 1986.

Turner, B.A., *Man-Made Disasters*, Wyeham, 1978.

Vanderplaats, G.N., and Blakely, K.D., "The Best and the Lightest", *Mechanical Engineering*, Volume 111, number 2, February 1989.

Van Dormolen, J., "Textual Analysis", in Christiansen, B., Hanson, A.G., and Otte, M. (Editors), *Perspectives on Mathematics Education*, D. Reidel, 1986.

Von Cranach, M., Kalbermatten, V., Indermohle, K., and Gugher, B., *Goal-Directed Action*, Academic Press, 1982.

Wallace, K.M., "Engineering Design Method", in Fourth International Conference on the Education and Training of Engineering Designers, Institute of Mechanical Engineers (London), 1985.

Watanabe, S., *Pattern Recognition: Human and Mechanical*, Wiley, 1985.

Wearne, S.H., "A Review of Reports of Failures", *Proceedings*, Volume 193, Institute of Mechanical Engineers (London), 1979.

Webber, J.H., "Keynote Address", 1986 Ship Production Symposium, Society of Naval Architects and Marine Engineers, 1986. Republished in *Marine Technology*, Volume 24, number 1, January 1987.

Webster, J.J., "Pulling - not Pushing - for Higher Productivity", *Mechanical Engineering*, Volume 110, number 4, April 1988.

Weck, R., "A Rational Approach to Standards for Welded Construction", *British Welding Research Association Bulletin*, Volume 7, 1966.

Weinberg, G.M., *An Introduction to General Systems Thinking*, Wiley, 1975.

Weiner, E.L., "Beyond the Sterile Cockpit", *Human Factors*, Volume 27, number 1, 1985.

Weinrib, J.M., "John Kenneth Galbraith: 'Beyond Technology Comes Good Design ...'", *Mechanical Engineering*, Volume 109, number 1, January 1987a.

Weinrib, J.M., "Richard Cyert: Managing to Get Into the Details", *Mechanical Engineering*, Volume 109, number 1, January 1987b.

Welter, T.R., "How to Build and Operate a Product-Design Team", *Industry Week*, Volume 239, number 8, April 16, 1990.

Wertheimer, M., *Productive Thinking*, Harper, 1959.

White, N.A., "Engineering Management - the Managerial Tasks of Engineers", *Proceedings*, Volume 197, number 68, Institute of Mechanical Engineers (London), 1983.

Wick, C., Benedict, J.T., and Veilleux, R.F. (Editors), *Tool and Manufacturing Engineers Handbook*, Volume II: Forming, Fourth Edition, Society of Manufacturing Engineers, 1984.

Willet, F., "The Case of the Derailed Disk Drives", *Mechanical Engineering*, Volume 110, number 1, January 1988.

Winograd, T., and Flores, C.F., *Understanding Computers and Cognition*, Ablex, 1986.

Winston, P., "Perspective", in Winston, P.H., and Prendergast, K.A., *The A.I. Business*, MIT Press, 1984.

Witwer, J.G., "An Interface for Innovation", *Mechanical Engineering*, Volume 111, number 6, June 1989.

Yeomans, R.W., Choundry, A., and ten Hagen, P.J.W. (Editors), *Design Rules for a CIM System*, North-Holland, 1985.

Zhang, Z., and Rice, S.L., "Conceptual Design: Perceiving the Pattern", *Mechanical Engineering*, Volume 111, number, 7, July 1989.

Zlatin, N., and Christopher, J., "Machining Characteristics of Difficult-to-Machine Materials", in Tipnis, V.A. (Editor), *Influence of Metallurgy on Machinability*, American Society for Metals, 1975.

Vita

Born in Chicago in 1956, Keith E. Schleiffer graduated in 1978 from the United States Coast Guard Academy with a Bachelor of Science degree in ocean engineering. For the next six years he served as an officer in the Coast Guard, as an Assistant Engineer on a ship and as a Project Officer in the Office of Research and Development. In 1983 he became registered as a Professional Engineer in the Commonwealth of Virginia. He is presently employed as a research engineer at the Battelle Memorial Institute.