# PERCEIVING, COMPREHENDING, AND MEASURING DESIGN ACTIVITY THROUGH THE QUESTIONS ASKED WHILE DESIGNING

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF MECHANICAL ENGINEERING

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

MECHANICAL ENGINEERING

Özgür Eris

August 2002

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Larry J. Leifer, Principal Adviser

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Bernard Roth

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Sheri D. Sheppard

Approved for the University Committee on Graduate Studies.

# Abstract

This work treats question asking while designing as a process, and examines its key aspects. The theoretical part of the research involves the development of a taxonomy of questions asked while designing, and an analytical framework for measuring design performance.

The contribution of the taxonomy—apart from proving to be a comprehensive analysis framework—is its ability to differentiate between Deep Reasoning Questions (DRQs), and Generative Design Questions (GDQs). DRQs reflect convergent thinking, whereas GDQs reflect divergent thinking. The contribution of the design performance framework is the distinction it makes between activity based internal (real-time) performance metrics and prototype based external performance metrics. Internal metrics are associated with the quality of the processes used to create the designs, whereas external metrics are associated with the quality of the designs.

The empirical part of the research involves designing and conducting experiments to test hypotheses generated from field observations. The more significant hypotheses postulate relationships between question asking processes of teams and their design processes, and between their combined DRQ+GDQ asking rates and performance. Both hypotheses were verified. Question asking processes of design teams were demonstrated to be descriptors of their design processes, and combined DRQ+GDQ asking rate was demonstrated to be an internal design performance metric.

The findings also demonstrated DRQ+GDQ utilization to be a mechanism designers rely on for managing divergent and convergent modes of thinking. During conceptualization, GDQs were shown to be instrumental in preserving ambiguity by reframing previously recognized needs and understandings, generating alternatives, and creatively negotiating proposed design concepts. During implementation and assessment, DRQs were shown to be instrumental in reducing ambiguity by reiterating goals, focusing on deliverables, seeking and establishing causality, and reducing the number of alternatives.

Special consideration was given to laying out the foundations of a unified design theory, which integrates the findings on question asking with existing understandings on decision making in design contexts. Application of the findings in developing better design information and knowledge creation and sharing systems was also considered.

# Acknowledgements

I consider myself rather fortunate to have joined the Design Division years ago as a Masters student. Over the years, I have grown increasingly cognizant and appreciative of the values that are promoted by its members.

Even though its sense of community cannot be solely attributed to specific individuals, I cannot help associating the two people, who have acted as my advisers from the very beginning, with its open, warm, jolly, exploratory, and yet, rigorous quality. I have learned so much from Larry and Bernie that I would not know how to pay them back. (And with Bernie, there are also all those dinners that I made him buy me in numerous, and unfortunately, rather unsuccessful, Turkish restaurants around the Bay area.)

Of course, there are many others, like Sheri, whose comments were very influential in improving the quality of this dissertation, like Rolf, whose insights always made me spin around and look at things from a different angle, and like Ade, who just about always had the time to listen and discuss.

I also consider myself rather lucky to have traveled to the United States at a young age, and to have been able to take advantage of the academic and personal opportunities that most likely did not exist elsewhere. I am grateful to my family, especially to my mother and father, who supported me with the difficult decision of leaving my native country, and with starting and leading a new life so far away.

I am also deeply grateful to all of my friends who were such good company throughout this effort, to Nisvan, Ergin, Vicki, Batu, Jim, Pete, Ken, Lawrence, Paul, Mark, Sian, Sam, Nese, and to the whole CDR crew!

# Table of Contents

# List of Figures

# List of Tables

# List of Transcript Excerpts

# 1 Perceiving, Comprehending and Measuring Design Activity through the Questions Asked while Designing: Overview

## 1.1 Motivation and Assumptions

Designing is question intensive. When compared to two other contexts for intellectual interaction, reading comprehension and classroom learning, designing promotes the asking of more and deeper questions. However, our knowledge of the role of question asking during designing is rather understudied and limited, and that is what I set out to explore in this dissertation. My main ambition is to gain the preliminary understanding design researchers currently lack on the topic.

The subject of question asking processes of design teams first attracted my attention during a video interaction analysis session aimed at hypothesis generation. The video data for the analysis were collected during a two week design project carried out by graduate engineering design students whose goal was to design, prototype and race a paper bicycle. During the analysis, I began to pay close attention to the questions raised in the interaction, and their effect on the design decisions that followed. Some questions seemed to have a strong effect on pivotal decisions, and others dissipated and had no discernable impact. In either case, questions and decisions struck me as being tightly linked at a conceptual as well as at a pragmatic level.

However, it quickly became clear to me that our understanding of questions—as they occur in a design context—was not comprehensive enough to allow me to study their relationship to other subjects such as decision making. I realized that I needed to know much more about the nature of questions that were asked while designing, and to be able to formalize certain aspects of their occurrence before I could relate them to the fundamental aspects of another subject. A review of the design research literature revealed little insight.

My reaction was to start work on the development of a theoretical framework on the nature of questions occurring in design contexts, and to apply it to quasi-controlled design situations for verification. However, I realized that by making a distinction between questions that are asked in non-design and design contexts, I was making crucial assumptions, and that they needed to be clarified.

Throughout this dissertation, I operate under two fundamental premises:

1) It is valid and useful to treat designing as a "way of thinking," and thus, as a specific type of cognition.
2) Question asking while designing is influential to the cognition of designers. It is related to the cognitive aspects of their problem solving, creativity, decision making, and learning processes, and consequently, to their overall performance.

### 1.1.1 Why Study Design Cognition?

For the most part, research in engineering is focused on understanding and predicting the behavior of artificial (man-made) systems by way of studying the physical principles that govern them. In practice, the fundamental competency of engineers is seen to be their ability to understand, synthesize and apply such natural principles in creating new technologies that ultimately result in new products.

There is no doubt that we, as engineers, benefit tremendously from studying and applying physical principles. However, as our knowledge of them has grown, it has become apparent that our personal involvement in the design process as human beings is also important, and that there is a need to understand the principles that govern *our* behavior as designers. One of the most intriguing components of that human dimension is related to the thought processes we use when we design; our thought processes—our cognition as designers—govern the behavior of the systems we design as much as the physical principles we apply to create them. Therefore, it is relevant to be concerned with what design cognition is, and how it can be studied and improved.

It is not clear when "design cognition" was first termed. In a keynote speech, Pahl presented a brief history of the collaboration between cognitive scientists and design engineers, and argued that the knowledge of technical systems was not sufficient in understanding the thought processes that led to the synthesis of designs, and that studying those thought processes was critical in improving the proposed design methodologies [Pahl 97]. Recently, several Ph.D. dissertations have been published as initial explorations in design cognition [Dylla 1991, Fricke 1993, Dorst 1997, Brereton 1999], and several research groups have began to address the topic directly [Gero 1990, Leifer 1992, Birkhofer 1995]. This activity indicates that design cognition is a prevalent approach that is attracting the attention of a growing number of engineering design researchers, and supports the first premise listed in the previous section.

## 1.2 Research Approach

This research consists of theoretical and empirical dimensions. The two dimensions are coupled, and build on each other; the results of the exploration in one dimension feed into and influence the exploration in the other dimension. The research principles that guided me throughout those explorations are summarized in the following sections.

### 1.2.1 Theoretical Dimension

The theoretical dimension of this research involves the development of two conceptual frameworks: a framework for categorizing the questions that are asked while designing, and another framework for measuring design performance.

#### 1.2.1.1 The Nature of Questions Asked while Designing

One way of studying the nature of questions that are asked while designing is to develop a comprehensive taxonomy of questions, and to utilize it as a coding scheme in analyzing the thinking of designers. When developing the taxonomy, there are various principles that can be used to differentiate[1] between the types of questions. For the purposes of this research, I focused on two such differentiating principles that are related: the conceptual meaning of questions, and a convergent—divergent thinking paradigm that is reflected in questions.

As I will discuss in detail in Chapter 2, the first principle has been articulated and used in the formulation of semantic question categories by Lehnert [Lehnert 1978]. Prior to adopting her categories and/or constructing any additional ones myself, I reviewed five other published taxonomies of questions. A common assumption reflected in their structure was that a specific answer, or a specific set of answers, *exist* for a given question. Also, two of the taxonomies seemed to be based on the assumption that the answers to questions are *known*.

When I considered those assumptions in light of my observations of design activity, and my own thinking as a designer, I felt that it was appropriate to propose a cognitive paradigm differentiating between convergent and divergent thinking modes. I considered the types of questions in the published taxonomies to be characteristic of *convergent* thinking, where the questioner is attempting to converge on "the facts," and is expecting the answer of his/her question to hold a truth-value.

On the other hand, the underlying assumption of the questions asked while designing is that they have, regardless of being true or false, multiple *alternative* known answers as well as multiple *unknown possible* answers. I considered such questions to be characteristic of *divergent* thinking, where the questioner is attempting to diverge away from the facts to the possibilities that can be generated from them.

The second principle of the taxonomy is based on this cognitive paradigm, and yields two high-level question categories under which the lower level question categories constructed through the application of the first principle can be placed. The

---

[1] The definition of a differentiating principle is taken to be an explicit rule, or a system of rules, that are used as the basis for expanding on a phenomenon and for constructing categories that fall under it. For instance, if physical appearance is taken as a differentiating principle for categorizing people, eye color, height, width and weight would constitute valid categories, whereas name would not be a valid category since it cannot be constructed through the application of the differentiating principle.

understanding embodied in these two principles resulted in the adoption of Lehnert's semantic categories, and in the formulation of new divergent question categories. Together, the categories formed a comprehensive taxonomy of questions that are asked while designing. The specifics of that framework will be discussed in Chapter 3.

#### 1.2.1.2 A New Perspective on Design Performance

The recognition of design cognition as a research topic in design research is beginning to influence our understanding of design performance. Traditionally, when considering design performance, researchers have been predominantly concerned with developing ways of evaluating the performance of the systems engineers design, and have been focusing on the outcome of the design process, the product. The recent focus on the human dimension of designing, and on design cognition, has introduced another perspective for considering design performance, the *designer*.

These two viewpoints suggest the existence of two types of design phenomena that can be evaluated: what occurs during design activity, and what results from and persists after design activity. Naturally, the metrics for evaluating the performance associated with each phenomenon will differ. If one grounds himself/herself in design activity, it is appropriate to treat activity-based metrics as being "internal," and to treat outcome-based metrics as being "external."

The significance and accuracy of the two types of design performance metrics depends on the context they are being used in. Since internal metrics focus on design activity, it is most appropriate to use them to judge the quality of the processes of design teams. And since external metrics focus on products of design activity, it is most appropriate to use them to judge the quality of the resulting designs.

As outlined in the second premise listed in the previous section, this research supposes the existence of a relationship between design cognition and performance. And since design cognition is a phenomenon internal to design activity, a framework for measuring internal design performance is required to study that relationship. When developing a framework in order to satisfy that requirement, I utilized the understanding resulting from the internal—external design performance distinction in formulating a question-centric internal design performance metric, and in relating that metric to design processes and cognition of design teams. The specifics of that framework will be discussed in Chapter 4.

### 1.2.2 Empirical Dimension: Three Experiments

The empirical dimension of this research entails carrying out of a series of detailed observations in two distinct settings, and analyzing the resulting data according to the two frameworks developed in the theoretical dimension. The first setting is a real-life design project, and lends itself to ethnographic observation techniques. The second setting is a quasi-controlled laboratory experiment, and lends itself to video interaction analysis. A detailed discussion on application of the observation and analysis techniques in these settings is provided in Chapter 4.

The research I conducted in the two settings can be summarized in three progressive steps:

1) Detailed observation and analysis of a real-life design situation for hypothesis generation.
2) Design of a laboratory experiment to test the hypotheses.
3) Evaluation and redesign of the pilot version of the experiment, and the execution of the final version.

In taking each step, I was influenced by Tang's design research methodology [Tang 1991], which advocates that the researcher should go beyond merely observing and describing design activity to constructing meaningful interventions to test the gained insights by iterating a cycle composed of three phases: observe, analyze, and intervene. The distinct nature of the three steps required me to put more emphasis on certain phases than others during each step [Figure 1-1].

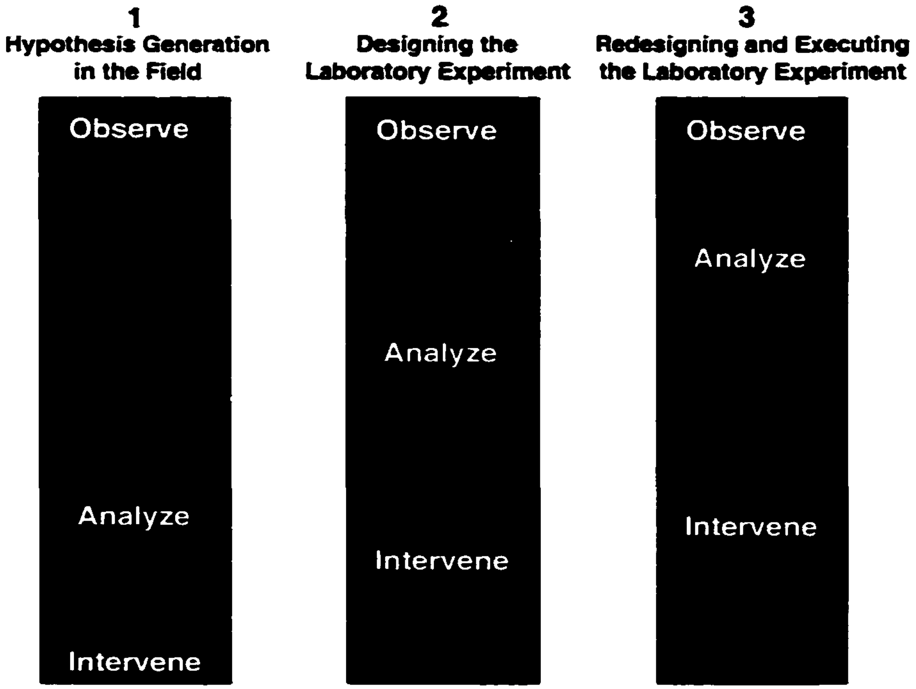


Figure 1-1. Tang's "Observe-Analyze-Intervene" cycle superimposed on the three steps of the empirical dimension of the dissertation. Each step entails multiple iterations of the cycle. Differences in the nature of the steps result in more emphasis on certain phases than other phases during each step. The relative sizes in the figure for each step are approximations for the time spent during each phase.

The structure associated with each empirical step is outlined in the following sections.

### 1.2.2.1 Hypothesis Generation in the Field

The first research setting, a real-life design project, enabled me to freely observe a design situation where a team of graduate engineering design students designed, prototyped, and raced a paper bicycle. A colleague and I "shadowed" the design team, videotaped the nine design meetings the team held over a period of two weeks, and carried out numerous video interaction analysis sessions after the project was over.

As mentioned earlier in this section, during those observations, I paid close attention to the questions raised in the interaction, considered potential relationships between question asking and decision making, and began to regard question asking while designing as a process. The pivotal research questions driving this dissertation stem from those initial observations and conceptualizations. A detailed discussion on those insights, and their development into testable hypotheses is provided in Chapter 4.

I generated 3 testable hypotheses from those observations:

**H1:** Question timing and type are descriptive characteristics of design cognition and process. When the set of questions a design team asks during a design project is considered as a whole, the timing and nature of those questions point at the fundamentals of the knowledge and rationale the team uses for breaking down and structuring the project into design phases. Question timing and type are informative enough to serve as a roadmap to the design thinking and process of the team.

**H2:** Overall question asking rate is related to design team performance, and can be taken as a design performance metric. There is a strong correlation between the frequency of questions and design team performance.

**H3:** Question asking behavior of design teams is influenced by their access to hardware. The types of questions design teams ask change when they transition from working in the absence of hardware to working with hardware.

#### 1.2.2.2 Design of the Laboratory Experiment

The second empirical step is the design of a laboratory experiment in order to test the hypothesis listed above. I identified seven design requirements under three experimental design criteria that needed to be met for the experiment to test the hypotheses. The framework for categorizing the questions that are asked while designing, the hypotheses, and experimental considerations specific to design research served as natural design criteria (the necessity of the framework for measuring design performance is implied by the first hypotheses related requirement listed below, R2.)

The specific requirements are the following:

##### Taxonomy Related Requirements

**R1:** The design experiment should promote realistic question asking processes from teams so that the application of the taxonomy of questions, which itself is based on data from realistic question asking processes, is meaningful.

##### Hypotheses Related Requirements

**R2:** Definitions and metrics for the phenomena outlined in the hypotheses should be developed prior to the execution of the design experiment.

**R3:** The design experiment should incorporate an intervention that promotes a clear distinction between designers working with and without hardware.

### Design Research Experimentation Related Requirements

**R4:** The design experiment should promote designing as opposed to problem solving.

**R5:** The setting and scenario of the design experiment should allow for the insertion of control elements associated with the hypotheses without overconstraining the activity (quasi-control as opposed to tight control).

**R6:** The design experiment should facilitate the testing of all hypotheses in a single experiment.

**R7:** The data collection methods used in the design experiment should result in data that can be analyzed qualitatively as well as quantitatively.

The nature of the requirements, and the specifications for meeting them, are discussed in detail in Chapter 5. However, I would like to stress the relevance, and even the necessity, of the third design criterion in design research experiments in general; the first three of the four requirements under it can be generalized.

In order to construct a design scenario that embodies the specifications, I identified an existing design exercise with known consequences, and modified it. In the exercise, the subjects were asked to design and prototype a measurement device called a "bodiometer," which can be moved along male and female body contours to measure their length, with an operating range from 3 to 100 inches. One group of teams, the control group, was provided with the prototyping materials at the beginning, and the other group of teams, the test group, approximately 35 minutes into the exercise.

### 1.2.2.3 Evaluation and Redesign of the Pilot Experiment: The Definition of a "Good" Question

In taking the third empirical step, I aimed to augment the hypotheses, and to ensure that the design exercise did indeed meet the experimental requirements. I conducted two pilot sessions of the experiment with six graduate mechanical engineering design students. The pilot runs proved to be very effective in achieving both goals.

Firstly, they resulted in changes to the structure of the design exercise and the application of the design performance metrics. Even though most of those changes were minor adjustments individually, their combined contribution to the improvement of meeting the requirements was significant. For example, observing a need to increase the duration of the exercise by 20 minutes during the pilot runs provided the teams in the final runs enough extra time to complete the number of design iterations they needed, which meant that R1 and R4 were met at a higher degree.

Secondly, the pilot runs allowed me to reflect on the relevance and validity of my hypotheses, and refine them as necessary. In that process, I was motivated to consider what a "good" question might be in a design context, and incorporate parts of its definition in my hypotheses by modifying H2 and creating H4.

The modification to H2 entails identifying two specific classes of questions that are hypothesized to be associated with design team performance, "Deep Reasoning Questions" (DRQs), and "Generative Design Questions" (GDQs). Graesser identified DRQs, and demonstrated a relationship between them and learning performance [Graesser 1988, 1993, 1994]. I identified GDQs while generating hypotheses and assessing the pilot runs, and hypothesized that DRQs and GDQs formed a complementary pair, which could serve as a performance metric in design contexts:

**H2:** Two classes of questions, termed Deep Reasoning and Generative Design questions, are related to design team performance. Their frequency of occurrence correlates strongly with design team performance, and can be taken as a performance metric.

I felt the need to create the new hypothesis, H4, when I considered the consequences of a "good" question in a design context as opposed to its definition. Revisiting my observations of the paper bicycle design team, I postulated that good questions were associated with, and yielded, conceptual leaps, or, discoveries.

**H4:** There is a strong correlation between the frequency of discoveries made by design teams and design team performance. Hence, discovery making can be taken as a performance metric.

I then conducted the redesigned version of the experiment with 36 graduate mechanical engineering design students working in 12 teams, analyzed the data according to the two theoretical frameworks, and tested the validity of the hypotheses. A detailed discussion on the redesign of the experiment and the modification of the hypotheses is provided in Chapter 6. The analysis of the data collected from the redesigned experiment is presented in Chapter 7.

## 1.3 Summary of Key Findings

The most significant finding that emerged from the theoretical dimension of this research is the extension of existing taxonomies of questions. The insights I gained on the nature of questions allowed me to identify five GDQ categories, and propose them as additions to the published taxonomies. The categories are: Proposal/Negotiation, Scenario Creation, Ideation, Method Generation and Enablement. Together with these additions, the extended taxonomy formed the basis of the analysis scheme for studying empirical data. It is presented together with the published taxonomies in Table 1-1.

| ARISTOTLE | DILLON | LEHNERT | GRAESSER | ERIS |
|---|---|---|---|---|
| Existence (Affirmation) | Existence/affirmation | Verification | Verification | Verification |
| | Instance/identification | | | |
| Nature (Essence/Def.) | Substance/definition | | Definition | Definition |
| | | | Example | Example |
| Fact (Attribute/ Description) | Character/description | Feature Specification | Feature Specification | Feature Specification |
| | | Concept Completion | Concept Completion | Concept Completion |
| | | Quantification | Quantification | Quantification |
| | Function/application | Goal Orientation | Goal Orientation ■ | Rationale/Function ■ |
| | Rationale/explication | | | |
| | Concomitance | Disjunctive | Disjunctive | Disjunctive |
| | Equivalence | | Comparison | Comparison |
| | Difference | | | |
| Reason (Cause/ Explanation) | Relation | | Interpretation | Interpretation ■ |
| | Correlation | | | |
| | Conditionality & Causality | Causal Antecedent | Causal Antecedent ■ | Causal Antecedent ■ |
| | | Causal Consequent | Causal Consequent■ | Causal Consequent ■ |
| | | Expectational | Expectational ■ | Expectational ■ |
| | | Procedural | Procedural ■ | Procedural ■ |
| | | Enablement | Enablement ■ | Enablement ■ |
| | | | | Proposal/Negotiation● |
| | | | | Enablement ● |
| | | | | Method Generation ● |
| | | | | Scenario Creation ● |
| | | | | Ideation ● |
| | | Judgmental | Judgmental | Judgmental |
| | Rhetorical | | Assertion | |
| | | Request | Request/Directive | Request |
| | Deliberation | | | |
| | Unspecified | | | |
| | Unclear | | | |

Table 1-1. A visual comparison of the categories of five taxonomies of questions. Dillon's categories are an expansion of Aristotle's. Graesser's and Eris's categories are an extension of Lehnert's. ■ denotes the types of questions termed as "Deep Reasoning Questions" by Graesser. ● denotes the types of questions termed as "Generative Design Questions" by Eris.

The empirical dimension of this research yielded several significant findings, and resulted in the validation of the first three hypotheses. Even though there was strong supporting evidence for the relevance of the fourth hypotheses, it could not be validated at a high enough significance level.

The most striking finding is the strength of the correlation hypothesized between combined DRQ+GDQ asking rate and design team performance (adjusted $R^2$ values of 0.68 for the control group, and 0.70 for the test group, with $p < 0.05$.) The data are plotted in Figure 1-2. On the other hand, there was no correlation between the asking rate of all types of questions, or of any single type of question, and design performance. Also, further analysis showed that DRQs and GDQs need to be treated as complementary pairs when it comes to establishing their value as a design performance metric.

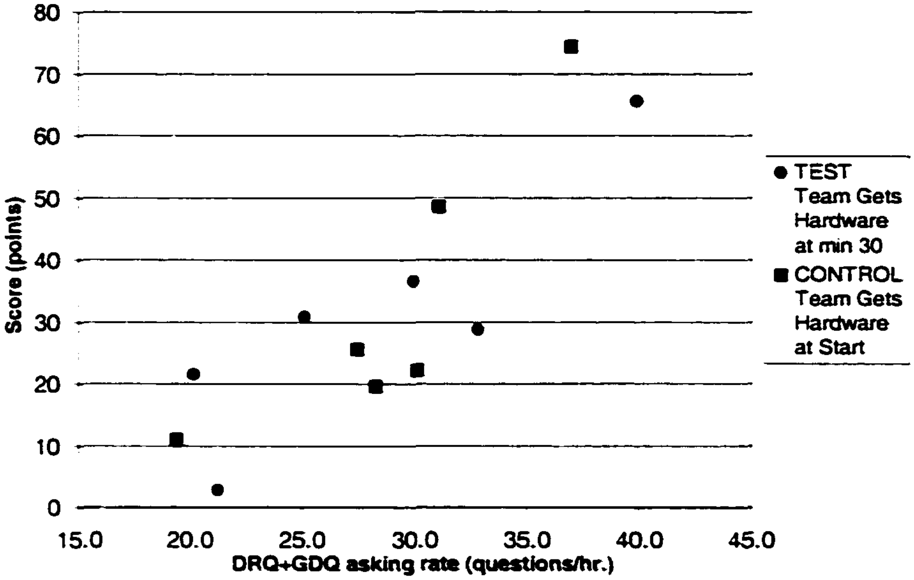


Figure 1-2. Combined DRQ+GDQ asking rates of the twelve design teams plotted against their prototype scores. Data points marked by squares belong to the teams in the control group. Data points marked by diamonds belong to the teams in the test group.

Considering the implication of the findings for the cognition of designers on a qualitative level allowed me to assign meaning to them by developing three views on the role of question asking in design. In Chapter 8, I synthesize the findings, and demonstrate that it is relevant, and beneficial, to treat question asking as:

1) A Process
2) Creative Negotiation
3) A Mechanism for Managing Convergent and Divergent Thinking Modes

Also in Chapter 8, I consider the verified hypothesis in conjunction with these three views, and draw the following conclusions:

1) Question asking reflects key aspects of design thinking and processes of teams. Furthermore, design thinking of teams evolves while question asking. While formulating questions—formulation of each question can be considered to be a

micro-design task—design teams find the opportunity to structure their design thinking by diverging and converging on their ideas.

2) The frameworks developed in Chapter 3 for characterizing and categorizing questions according to their conceptual meaning, and in Chapter 4 for measuring design performance are valid, and have potential for further development.
3) The question-based metric derived in this study not only measures design performance, but also serves as a descriptive "lens" for revealing and monitoring the thinking of designers during design activity.
4) Question asking, hence design thinking, of teams is strongly influenced by their access to hardware. When conceptualizing in the absence of hardware, design teams exhibit more divergence in their thinking by relying more on Generative Design Questions.

## 1.4 Guide to the Dissertation

Chapter 2 presents reviews of the design research field and published taxonomies of questions. Current research areas are categorized into four topics, and design cognition is positioned within them. Also in Chapter 2, relationships between two fundamental cognitive mechanisms in designing, decision making and question asking, are proposed.

Chapter 3 describes the development of a comprehensive framework for categorizing the questions that are asked while designing. The rationale behind constructing 5 new question categories is discussed in detail.

Chapter 4 presents an overview of the three steps of the empirical dimension, and describes the first step, hypothesis generation in the field. Also in Chapter 4, a framework for measuring internal design performance is presented.

Chapter 5 describes the second empirical step, the design of a laboratory experiment to test the hypotheses. The requirements for the experiment, and the specifications for meeting them are discussed in detail.

Chapter 6 describes the redesign of the experiment and the modification of the hypotheses.

Chapter 7 presents the analysis of the data collected from the redesigned experiment in order to test the hypotheses.

Chapter 8 presents the implications of the findings and the conclusions. Also in Chapter 8, the contributions of this dissertation to design research are listed, and opportunities for future research are discussed.

# 2 Question Asking: A Fundamental Cognitive Dimension

As mentioned in Chapter 1, I operate under two premises throughout this dissertation:

1) It is valid and useful to frame designing as a "way of thinking", and thus, as a specific type of cognition.
2) Question asking while designing is influential to the cognition of designers. It is related to the cognitive aspects of their problem solving, creativity, decision making, and learning processes, and, consequently, to their overall performance.

These premises have two major implications. The first implication is that studying design cognition is a distinct and relevant approach to design research. The second one is that treating decision making as the fundamental cognitive mechanism driving design performance—a prominent position within the field—requires further consideration.

This chapter consists of three parts. The first two parts, Sections 2.1 and 2.2, stem from my motivation to put those implications into perspective. Section 2.1 deals with the first implication, and entails reviewing the design research field by categorizing the current research areas into four topics, and positioning design cognition within them. Section 2.2 deals with the second implication, and entails focusing on design cognition by proposing and considering relationships between two fundamental cognitive mechanisms in designing, decision making and question asking.

The third part, Section 2.3, is a review of published taxonomies of questions. It represents my initial exploration on the nature of questions, and constitutes the first step in the synthesis of a coding scheme that can be used to analyze the question asking processes of designers.

## 2.1 Contemporary Topics in Design Research

In the next four sections, I put the first implication listed at the beginning of this chapter into perspective by discussing the contemporary topics in design research and positioning design cognition within them. I classify the topics into four categories: design processes, social theories, design information, and design cognition.

After an initial consideration, one might argue that the four categories I propose overlap to the degree that they lack meaning. The categories are indeed strongly related. Nevertheless, I see them as being defined by well-pronounced differentiations within the field, strongly reflected in the motivations and products of distinct groups of researchers. On the other hand, I believe that the strong relationships, and even overlaps, between the topics can and should act as a basis for informing researchers on the missing key aspects of knowledge within their domains. For example, much of the lacking functionality in the systems design information and knowledge support researchers develop can be alleviated by utilizing the findings from the other three domains—it is poor practice to develop a design knowledge sharing tool which does not address the underlying social, cognitive and process related elements of designing.

### 2.1.1 Design Processes

Researchers studying design processes have traditionally been concerned with categorizing the workflow of designing by braking it down to specific interrelated tasks. Their main goal is to formulize processes of designing, and to derive methods for design practice from them.

Numerous design processes have been developed by different researchers [Asimov 1962, Hubka 1982, Pugh 1986, Pahl & Beitz 1988, Ullman 1992, Otto & Wood 2001]. Since processes are abstractions—in most cases, derived from observation—the principles for abstraction can and often do differ between researchers. However, the basic tasks that make up processes are generally shared. What differentiate them are the specifics of the relationships between the tasks and procedures they embody.

A representative and influential design process model developed by Hubka is illustrated in Figure 2-1 [Hubka 1982]. It outlines tasks and procedures for designing. Arrows pointing back at prior tasks indicate iteration procedures.

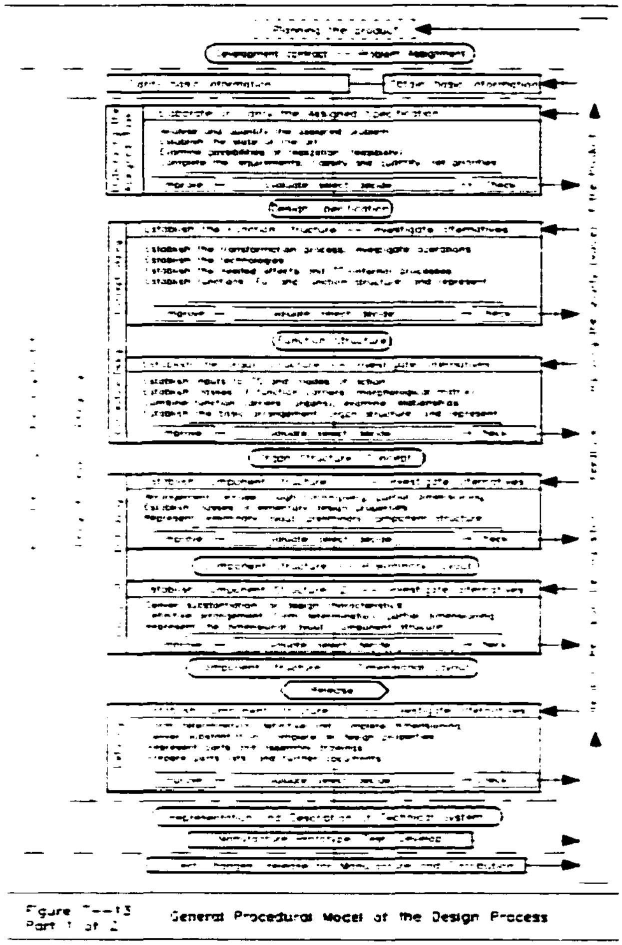


Figure 2-1. A design process model outlining tasks and procedures [Hubka 1982].

Another prominent design process, developed by Pahl and Beitz, is illustrated in Figure 2-2 [Pahl & Beitz 1988]. Upon its introduction, it has been recognized as an official standard in Germany and has been widely applied in industry for designing new products.

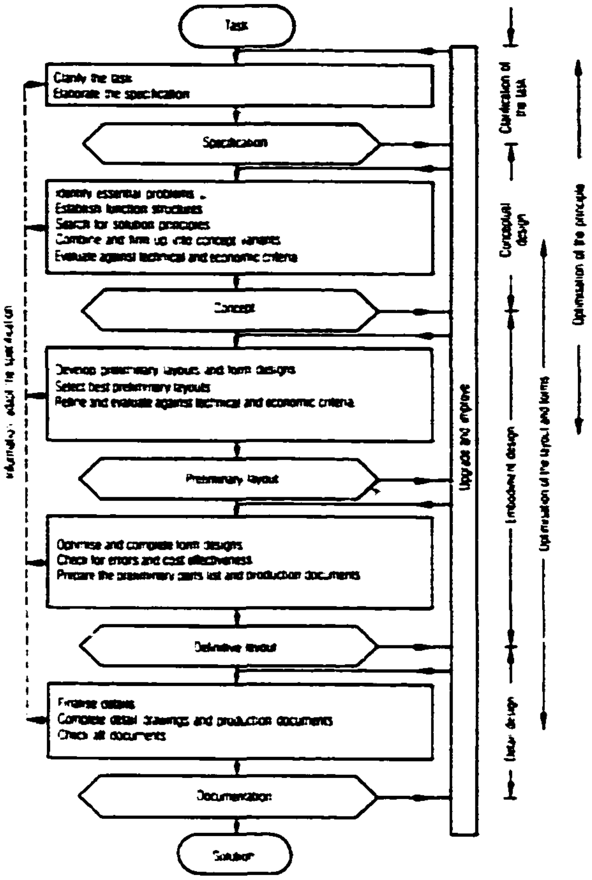


Figure 2-2. An influencial design process model—a standard in German industry [Pahl & Beitz 1988].

The tasks that serve as the basic elements for the two models are indeed similar; both processes address tasks related to the generation of requirements, concepts, specifications, layouts, and representations under one name or another. However, they propose somewhat different procedures for executing them.

There are two domains for considering the practice and utility of design processes: institutions, and individual or small groups of designers. For institutions, design processes constitute directly applicable methods that can be used to structure product development projects. They also constitute the basis for organizing the social and physical elements of most modern product development institutions; a group of people and space are associated with each task, i.e. requirements engineers, release engineers, test engineers, concept development laboratories, testing facilities, and manufacturing plants, etc. In other words, in institutional settings, design processes have social and physical manifestations.

For individual or a small group of designers, design processes constitute methods that can be internalized and practiced while designing. It is reasonable to assume that they affect the way designers think (this relationship will be discussed in detail in Section 4.4.4). In order to determine if that is the case, one needs to observe how designers communicate and act since it is difficult to directly observe how they think. In other words, design processes do not necessarily have physical manifestations in the practices of individual designers.

### 2.1.2 Social Theories of Design

Social theories of design are essentially constructionist approaches. Researchers who are interested in developing social theories aim to describe design activity by observing, analyzing and reconstructing the social interactions of the involved parties. They primarily focus on the social elements of designing (the effects of the internal social relationships on the activity itself and its outcomes) rather than the social implications of designs (the effects of the outcomes of the activity on external social relationships).

One of the earliest researchers who studied designing with that framework is Cuff. Her work focuses on the negotiation that takes place between architects and clients in design practice [Cuff 1982]. She challenges the myth of the architect being the driving factor in architectural design. She argues that, in practice, influence is "diffused" across all participants, including clients, and that qualities such as ambiguity, unexpected outcomes and open-endedness are inherent elements of designing. She concludes by stating that the final design *emerges* out of the interaction of the participants.

Bucciarelli studied two engineering design projects in industry by utilizing ethnographic methods [Bucciarelli 1988, 1994]. The main premise for his study is consistent with the main conclusion of Cuff: design is a social process. Bucciarelli acknowledges the pivotal role of the social interaction in design, and goes further by observing that:

> "Different participants think about the work on design in quite different ways. They do not share fully congruent internal representations of the design."

He uses that observation to propose the existence of "object worlds," which are "worlds of technical specializations, with their own dialects, systems of symbols, metaphors and models, instruments and craft sensitivities." In essence, he argues for the existence of a relatively engraved set of internal values that are inherent to each participant. The values act as filters when the participants perceive and relate to others. For example, when working together, a structural engineer will relate to the design activity by focusing on the strength of the design whereas a manufacturing engineer will do so by focusing on the part count and complexity of the design. Even though they look at the same thing, their mindsets determine their viewpoints, and they see different things. Building on that observation, Bucciarelli argues that the resulting design is not simply a summation of the products of those viewpoints, but rather, that it is their *intersection*.

Minneman studies an engineering design team and a series of design exercises that took place in a workshop [Minneman 1991]. He advocates going beyond mere observation, to intervention, in order to test the gained insights. His initial findings reiterate Cuff's conclusions. He also reemphasizes Cuff's and Bucciarelli's views on the role of ambiguity and negotiation, that they are inherent to designing and constitute a condition and a mechanism for understanding and structuring design activity. His contributions come in the form of implications. He argues that:

- "Those insights[2] shift the focus of group design support onto communication systems."
- "Design education should be refocused on teaching designers to better function in group situations."
- "Design management must encourage designers to work together."

And finally, it should be noted that the synergistic contributions of these three studies encouraged further interdisciplinary approaches to design research by demonstrating

---

[2] Insights on the role of ambiguity and negotiation in designing.

value in the application of cross-disciplinary analysis frameworks and methods to engineering practices.

### 2.1.3 Design Information

Researchers concerned with understanding the generation, capture and sharing of design information are heavily influenced by the recent developments in digital information technology. Even though the term "information" is not formally defined in most of their publications [Eris 1999], there seems to be an informal understanding of what it represents. The following is the closest definition I can construct to that understanding: design information is the content of communication generated while designing which needs to be contextualized in order to gain meaning.

The researchers' treatment of information leads me to associate information with communication in this definition. There seems to be a similarity in their usage of the word information[3], suggesting that, in a design context, all information is created with the intent of being communicated—if not right away, sometime in the future. Their usage also leads me to view information as lacking any specific meaning; the communication needs to be interpreted for it to be assigned meaning, in which case it might be more appropriate to call it knowledge.

When studying design information, most researchers aim to implement their findings in software tools that support information communication, capture and reuse. The requirements for such systems are usually based on findings on the information-handling behavior of designers obtained through observation.

Kuffner and Baya are two of the researchers who directly focus on understanding the information-handling behavior of designers during conceptual design [Kuffner 1990, Baya 1996]. Kuffner's framework is based on the formulization of the information requests of designers while they design. He pays special attention to "the design information required to answer questions about the design and to verify, and refute conjectures about the design." His contribution is to demonstrate that designers are

---

[3] For instance, the usage by McMahon and Wood [McMahon 1999, Wood 1999].

interested in information other than that which is contained in traditional design documentation such as blueprints and specifications. He also briefly discusses the possibility of developing a software tool that supplies the design information that is not contained in traditional documentation.

Baya utilizes a similar approach, and in a preliminary study, explores the question asking behavior of designers in order to understand their information needs. He goes one step further than Kuffner by incorporating his initial findings into the development of an information management tool, DEDAL. The deployment and assessment of DEDAL in design situations enables him to obtain some key results regarding the information-handling behavior of designers. He discovers that designers move between different types of information on an average of 13 seconds, and that they handle up to 40 concepts at a time while they design.

In light of such empirical data, Yen argues that concept generation and development occur most frequently in informal media where capture tools are the weakest, and develops a software tool, RECALL, that captures tacit information generated in multimodal design activity [Yen 2000]. By deploying RECALL, he demonstrates that the capture and playback/analysis of tacit information during concept development reveals the rationale behind the decisions that were made.

Yang anticipates the growing role of electronic information in design activity, and aims to enhance the collaboration among design teams by developing a software tool that improves the indexing and retrieval of design information [Yang 2000]. Similar to Yen, she perceives value in capturing and indexing design information while it is being generated. Making the analogy to a traditional engineering logbook, she qualifies her tool as an "electronic notebook," and argues that it provides a "rich, unfiltered history of a design project."

Frankenberger takes a different position; based on observations of engineering design processes in industry, she argues that it is more revealing to study the information-handling behavior of designers with respect to the design situations they are in [Frankenberger 1999]. She distinguishes between routine work and critical situations, and reports that designers contact their colleagues for information in nearly 90% of the critical situations. She argues that the information needs of designers can be adequately supported by software tools only during routine work, and that during critical situations, social interaction cannot and should not be substituted for.

### 2.1.4 Design Cognition

The topic this dissertation falls under, design cognition, involves the study of the thought processes designers experience while they design. It might be appropriate to refer to such thought processes as *design thinking*; since cognition is defined as "the act of knowing,"[4] it is plausible to treat design cognition as being synonymous with design thinking.

Researchers who study design cognition focus mainly on the individual designer. That attribute of design cognition differentiates it from the other contemporary design research topics. Studying the other topics involves focusing on mechanisms and relationships that can be considered to be external to the individual designer, i.e. design tasks and procedures, information flow, social interaction. That is not to say that, under design cognition, the individual designer is treated as an isolated entity whose internal mechanisms have little connection with other designers or the environment. On the contrary, studying such relationships might prove to be useful in discovering what is taking place *inside* the individual designer's mind. Brereton's dissertation is a good example of such a framework, where the interactions between designers and hardware are treated as constituents of "distributed cognition" [Brereton 1999], and uses them to understand the development of the individual designer's cognition and learning processes while designing.

Most studies on design cognition are a direct result of the application of cognitive science theories and methodologies that psychologists and artificial intelligence researchers have developed to explaining and modeling design activity. Lehnert, an artificial intelligence researcher, writes [Lehnert 1978]:

---

[4] As defined in the Longman Contemporary Dictionary of English.

> "Among scientists interested in cognition, there is no general agreement on how it can be best studied. Cognitive science is therefore characterized as an interdisciplinary area, to which contributions may be made by either computer scientists or psychologists. This may seem surprising at first, since computer science and psychology are not commonly considered strongly related fields of interest. Once one understands exactly how a computer scientist and a psychologist go about studying cognitive phenomena, however, the connection is less mysterious."

She then outlines the research methodologies of psychologists and computer scientists, compares them, and concludes their frameworks are analogous other than the fact that psychologists conduct experiments and computer scientists write programs. Her point is that both are useful paradigms for testing educated guesses. She sees the two paradigms as being complementary since some cognitive behavior can be studied better with experiments, and others, with computer programs.

Although Lehnert's views are more than twenty years old, they still have value. The distinction she makes between the experimental and computational research methods utilized in discovering cognitive behavior is visible even today in design research: some design researchers study design cognition by building computational models of designer behavior, such as Gero [Gero 1985], and others study it by designing experiments simulating design situations with real designers, such Cross and Dorst [Cross 1996]. This dissertation falls under the second paradigm.

Theoretical research methods constitute a third paradigm. They aim to model the cognition of individual designers based on personal experiences and first hand observations, and therefore, are more subjective when compared to the other two methods. A good example of a researcher who employs a theoretical approach is Schon. In his influential work The Reflective Practitioner, he proposes a framework that describes the individual designer's "professional artistry," which consists of five basic elements: knowing in action, reflection in action, conversation with the situation, reflecting on the situation, and reflective conversation with the situation [Schon 1983]. His framework is not meant to be an objective model that can be tested and verified scientifically. It is meant to be an intellectual paradigm for discussion as well as self-reflection.

## 2.2 Question Asking as a Fundamental Dimension of Design Cognition

Within the design cognition domain, much has been published on the roles of learning, knowledge sharing and management, visualization, problem solving, and decision making in designing. These subjects have all been formulized to a degree in disciplines outside of design research. Design researchers have been making contributions by applying those understandings to describing and modeling design activity. However, I believe that a fundamental cognitive dimension, question asking, has been omitted. This is most likely the result of the absence of a unified and process-oriented theory of question asking within as well as outside of the design research field.

Therefore, in this section, I set out to demonstrate the significance of question asking as a cognitive mechanism in designing. I intend to accomplish that by supporting the validity of the implication of the second premise I listed at the beginning of this chapter (treating decision making as the fundamental cognitive mechanism driving design performance requires further consideration) by discussing three prominent decision-centric views in design research, and by arguing for an inherent duality between question asking and decision making.

### 2.2.1 Decision Making as a Cognitive Mechanism which Drives Design Performance

Recently, an increasing number of researchers have been arguing for decision-centric models of engineering design, which treat decision making as a fundamental cognitive mechanism that drives performance [Hazelrigg 1999, Gero 1985]. Hazelrigg offers the following view:

> "It order to ensure that engineering design is conducted as a rational process producing the best possible results given the context of the activity, a mathematics of design is needed. It is possible to develop such a mathematics based on the recognition that engineering design is a decision-intensive process and adapting theories from other fields such as economics and decision theory."

He builds on that argument by utilizing decision theories in constructing a set of axioms for designing, and in deriving two theorems. He illustrates his view by considering a scenario where several people are attempting to guess the number of M&Ms in a jar. He first tackles the scenario through what he calls the "conventional engineering approach", which entails modeling the volumes of the jar and individual M&Ms and relating them to each other. He then tackles it by applying his theorems in producing a statistical model, which accounts for uncertainty, risk, information, preferences, and external factors such as competition (elements of Game Theory). His model results in a number of decisions one of which he computes as being optimal. He represents his approach in the form of a decision tree (Figure 2-3).

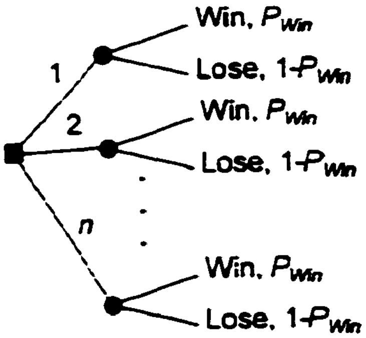


Figure 2-3. Decision tree for competitively guessing the number of M&M in a jar [Hazelrigg 1999].

He compares the traditional approach with his, and concludes that his axiomatic approach yields a more accurate representation, and produces results with a higher probability of winning. In his closing words, he remarks that "all engineering design is a matter of decision making under uncertainty and risk."

Radford and Gero also hold a decision-centric view [Radford & Gero, 1985]. Their goal is similar to Hazelrigg's as both parties are interested in constructing mathematical models of designing. Their approach differs when the nature of their models is considered; Radford and Gero advocate constructing deterministic models and account for dealing with ambiguity through optimization, whereas Hazelrigg advocates constructing probabilistic models which have elements of ambiguity already built in.

Radford and Gero begin by acknowledging that design activity can be approached in different ways and that many paradigms—numerical and qualitative—exist for understanding it, and provide their rationale for taking a decision-centric view:

> "As a starting point we shall take the premise that the essential feature of design is the existence of goals—however ill-defined those goals—which makes the process purposeful and necessitates decisions about the best way to achieve those goals."

Then, they argue for a relationship between such decisions and the performance of the solutions they lead to, and state the following:

> "The exploration of the relationships between design decisions and solution performances is fundemental to design—a process of predicting the performance consequences of design decisions and postulating the decisions which will lead to desired performance resultants."

In order to formulize that process, they argue for the existence of three mathematical paradigms: simulation, generation, and optimization. They see optimization as "building on both the simulation and generation paradigms" and "introducing goal-seeking directly into the process."

Dieter is more pragmatic; he is concerned with design practice. He demonstrates the relevance of applying existing decision-centric views in evaluating and choosing between alternative design concepts [Dieter 1983]. After briefly discussing decision making under risk and uncertainty, he illustrates the construction of a decision matrix in order to determine the utility values—intrinsic worth of outcomes—associated with competing design concepts (Figure 2-4). His method is based on utility theory, which formalizes the development values in decision making.

| Objective | Weight factor | Parameter | Built-up plates, welded |  |  | Built-up plates, riveted |  |  | Cast hook |  |  |
|---|---|---|---|---|---|---|---|---|---|---|---|
|  |  |  | Magnitude | Score | Value | Magnitude | Score | Value | Magnitude | Score | Value |
| Material cost | 0.10 | ¢/lb | 25 | 8 | 0.8 | 25 | 8 | 0.8 | 30 | 9 | 0.9 |
| Manufacturing cost | 0.20 | $ | 1500 | 7 | 1.4 | 1200 | 9 | 1.8 | 2000 | 4 | 0.8 |
| Time to produce | 0.05 | hours | 40 | 7 | 0.3 | 25 | 9 | 0.4 | 60 | 5 | 0.2 |
| Durability | 0.15 | experience | high | 8 | 1.2 | high | 8 | 1.2 | good | 6 | 0.9 |
| Reliability | 0.30 | experience | good | 7 | 2.1 | excellent | 9 | 2.7 | fair | 5 | 1.5 |
| Reparability | 0.20 | experience | good | 7 | 1.4 | very good | 8 | 1.6 | fair | 5 | 1.0 |
| Overall utility value |  |  |  |  | 7.2 |  |  | 8.5 |  |  | 5.3 |

Figure 2-4. A decision matrix used to determine the utility values associated with competing design concepts [Dieter 1983].

He then introduces probability theory, which assesses the states of knowledge, and combines them with elements from utility theory in demonstrating the application of decision trees. Even though the example he uses, management of an R&D project, does not seem to be directly relevant, when viewed from a process point of view, it can be used to model certain aspects of design activity (Figure 2-5).

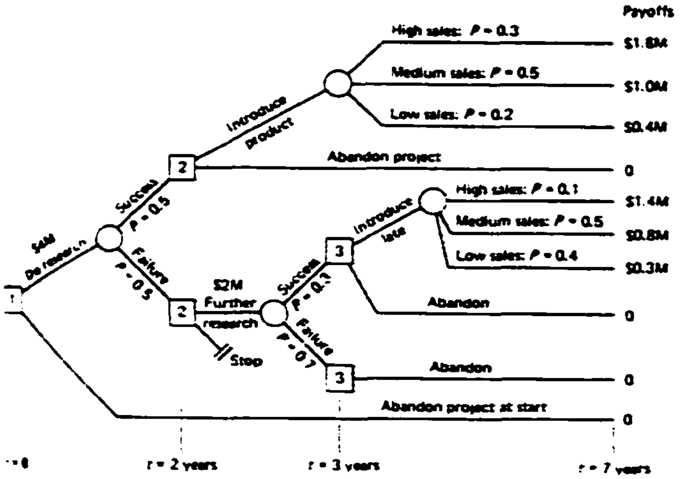


Figure 2-5. Decision tree for the management of an R&D project [Dieter 1983]. Squares indicate decision points (in control of the decision maker) and circles indicate chance events (out of control of the decision maker).

As Hazelrigg, Radford and Gero, and Dieter have all argued for, designers are faced with crucial challenging decisions after generating concepts, which constitute different choices with different outcomes. Applying decision theory principles can improve their decision making processes by aiding them in choosing the most appropriate concept to satisfy a certain set of constraints, preferences and goals. However, there are limitations to modeling designing as a decision making process, and to treating design making as a cognitive mechanism that drives design performance. Since designing is human and complex, there must surely be many other cognitive dimensions that drive design performance. Current decision-centric views would benefit from the consideration of potential relationships between decision making and other cognitive mechanisms used while designing.

### 2.2.2 Associating Questioning Asking and Decision Making: Two Axiomatic Interdependencies

Studying decision making as a rational process, and considering its role in designing is valuable. The validity of studying decision making as a rational process does not need explicit qualification as it has been rigorously argued for in many different domains. As Howard remarks, decision analysis is related to "the systematic reasoning about human action," and it "stands on a foundation of hundreds years of philosophical and practical thought" [Howard 1988]. He states that the "resurgence of the field in modern times began with statistical decision theory and a new appreciation of the Bayesian viewpoint." His definition of the modern discipline of decision analysis is "a systematic procedure for transforming opaque decision problems into transparent decision problems by a sequence of transparent steps."

I touched upon the role of decision making in designing in the previous section by suggesting that the relationships between decision making and other cognitive mechanisms used while designing need to be considered. I believe the most effective way of addressing that issue is simply to ground the motivation and context of decision-centric studies of design in observations of design activity.

I have been interested in the role of decisions in designing myself. As I mentioned in Chapter 1, while I was observing a design team in the field, I was motivated to pay close attention to the questions raised in the interaction, and their effect on the design decisions that followed. Some questions seemed to have a strong effect on pivotal decisions, and others dissipated and had no discernable impact. In either case, questions and decisions struck me as being tightly linked at a conceptual as well as a pragmatic level. I became interested in exploring that connection by marking all observable questions and decisions that occurred in the design team interaction and constructing a question-decision map. My intent was to determine if such a representation might be useful in confirming the existence of a connection and in articulating relationships between the nature and timing of questions and the decisions they lead to.

When I attempted to construct a representation, I quickly realized that our understanding of questions—as they occur in a design context—was not comprehensive enough to allow me to study their relationship to other subjects such as decision making. Therefore, as the initial step, I decided to focus on the question asking processes of designers in this dissertation, and to tackle the question-decision relationship in the future. However, I also decided to elaborate on what that relationship might be, and to treat that initial conceptualization as a vision in order to remain connected with its significance.

As a designer who is keen on appreciating the design cycle as a whole, when faced with decision centric approaches to designing such as the ones discussed in the previous section, and especially with decision tree structures that outline and associate information and knowledge regarding decisions with a decision process, my instinctive reaction is to ask the following two questions:

1) How did the decision-maker reach a position where he/she realized he/she could formulize what he/she knew into that structure?
2) How is reaching that position related to the decision making process, and more importantly, to the design process as a whole?

Design researchers tend not to consider those issues. That can lead to treating the decision making process *as* the design process, which, in my view, is an unsound analogy. On the other hand, decision theorists acknowledge those issues. However, their main focus remains on the decision making process. What they rigorously formulize and construct as a science of decision analysis seems to take place after that position is reached. For instance, Howard asks, "Is decision analysis too narrow for the richness of the human decision?" He then argues that issues such as "framing" and "creating alternatives" should be seriously considered before decision analysis techniques are practiced to ensure that "we are working on the right problem" and that "we can deal with problems without well defined alternatives" [Howard 1988]. More specifically, with regard to framing, he writes:

> "Framing is the most difficult part of the decision analysis process; it seems to require an understanding that is uniquely human. Framing poses the greatest challenge to the automation of decision analysis."

Even though he seems to consider framing to be an initial stage of the decision analysis process, he does not provide the type of systematic normative methods he proposes for the latter stages of the process for dealing with it.

The activities Howard identifies as being problematic, framing and creating alternatives, are inherent components of designing. Design researchers have been attempting to formulize them for decades as core elements of their design theories. Therefore, it would be sound to claim that while design researchers have much to learn from decision theorists, decision theorists have also much to learn from design researchers. In other words, it would be productive to unite our knowledge on decision making with our knowledge on designing.

In light of this discussion, let us return to the first question I posed earlier, "How did the decision maker reach a position where he realized he could formulize what he knew into a decision process structure?" Two ways of attempting to answer it can be: to claim that we get there via intuition, and therefore do not have or need a formal understanding of it, or to attempt to borrow from existing design theories. However, as I mentioned earlier, design researchers do not to formulize the relationships between decision making and other key aspects of designing when constructing design theories, so there is not much that can be borrowed and applied directly.

Here, I propose a third way of answering that question: to ask another question and let its answer guide me to the existence of a duality between question asking and decision making. The question is this: "How back-derivable is a decision making process?" Or in other words, "If one was to start with a decision and work his way back through all the cognitive events that led to that decision, what would he do when he reached junctions in the decision process which were associated with clusters of information and knowledge?"[5]. As an answer, I propose that one needs to consider the *questions* that made the construction of those clusters of information and knowledge possible, which means that one needs to analyze and understand the question asking processes of the decision maker.

I will illustrate that view with the following data segment extracted from one of the experiments I conducted in the empirical dimension of this research. In the experiments, teams of 3 graduate mechanical engineering students were asked to design and prototype a device that measures the length of body contours. In this specific excerpt, the team members are trying to decide on the number of the gear reduction stages between the sensor and the readout of the device in order to provide a meaningful measurement to the user (Transcript 2-1). Under the far right column, 14 questions and a decision that occur during the interaction are tagged sequentially.

---

[5] This specific formulation was introduced to me by Larry Leifer during a private discussion in 2000.

| In | Out | Voice | Utterance | Tag |
|---|---|---|---|---|
| 5:04:13 | 5:04:15 | A | So, what kind of gear reduction did we decided we needed? (No decision has been made) | Q1 |
| 5:04:18 | 5:04:21 | C | So, 0.25 inches... | |
| 5:04:22 | 5:04:24 | B | the circumference is... | |
| 5:04:25 | 5:04:26 | C | 7...4...5... | |
| 5:04:27 | 5:04:29 | B | Do we wanna know the circumference then? | Q2 |
| 5:04:32 | 5:04:32 | C | Right, not the area. | |
| 5:04:33 | 5:04:35 | B | The circumference is 2 Pi R? | Q3 |
| 5:04:36 | 5:04:36 | A | Yep. | |
| | | | ...(team calculates circumference together)... | |
| 5:05:12 | 5:05:14 | B | So we want something to only go around once? | Q4 |
| 5:05:17 | 5:05:18 | C | Right, 50 revolutions. | |
| 5:05:21 | 5:05:21 | B | 150? | Q5 |
| 5:05:24 | 5:05:25 | C | Right. How many teeth are on these guys (gears)? This one has 5,6,7,8. | Q6 |
| 5:05:29 | 5:05:33 | A | Or we can also do the belts. We can have rubber bands, yah. | |
| 5:05:39 | 5:05:40 | A | Can I borrow the ruler? | |
| 5:05:42 | 5.05:45 | B | It seems like there are...Oh, it says on them actually. 24. | |
| 5:05:47 | 5:05:49 | C | That's 3. 3 to 1. | |
| 5:05:52 | 5:05:53 | B | And we need 50 to 1? | Q7 |
| 5:05:54 | 5:05:54 | C | Yep. | |
| 5:05:55 | 5:05:55 | B | Hmmm. | |
| 5:06:03 | 5:06:07 | A | This is about a quarter of an inch, three quarters of an inch. (measuring with ruler) | |
| 5:06:08 | 5:06:09 | C | So, we'd actually need 3 stages? Is that right? | Q8 |
| 5:06:16 | 5:06:16 | B | 3 times 3 to the 2 is 27... | |
| 5:06:19 | 5:06:21 | C | So that would still give us 2 revolutions. | |
| 5:06:22 | 5:06:24 | B | Yeah, we need at least 4 stages. | |
| 5:06:30 | 5:06:32 | C | That should be kind of hard to read, wouldn't it? | Q9 |
| 5:06:36 | 5:06:46 | A | Well, maybe we can rotate around twice? I mean it's not hard to realize if it rotates around once, then we just need to aim for half of that. Do you know what I mean...maybe... | Q10 |
| 5:06:47 | 5:06:48 | C | So, which one of you has the smaller hands? | Q11 |
| 5:06:49 | 5:06:52 | A | I have the smaller, probably smaller. I have long fingers. | |
| 5:06:54 | 5:06:55 | B | What was, what were yours? | Q12 |
| 5:06:57 | 5:06:57 | C | 40 inches. | |
| 5:06:58 | 5:06:58 | B | 40 inches... | |
| 5:07:01 | 5:07:09 | C | So, with the smaller hand if you go around, and if it's over 27 then it doesn't matter if it goes around more than once. | |
| 5:07:09 | 5:07:20 | A | I would say that after we could have it go...the indicator could rotate around twice and a little bit before it's hard to read. Do you know what I mean? | |
| 5:07:21 | 5:07:24 | C | Okay, 3 stages seems appropriate, right? | Q13 |
| 5:07:25 | 5:07:25 | B | Yes. | D1 |
| 5:07:27 | 5:07:29 | A | Is that assuming that we have a bunch of little gears though? | Q14 |
| 5:07:31 | 5:07:39 | C | I'm kind of going under the assumption that we'll get about the same the gear ratio out of the rubber bands, too, since they're about the same size. | |

Transcript 2-1. Design team members A, B, and C are making a decision on the number of stages of gear reduction between the sensor and the readout so that their device provides a meaningful measurement to the user. Under the far right column, 14 questions and 1 decision that occur during the interaction are tagged sequentially.

The most striking observation is that all 14 questions are directly related to the decision the team is considering, and influence the 3.5 minute process that leads the team to a consensus by providing structure for the discussion and generating/uncovering the necessary information.

The process is initiated by A, who brings up the issue of the need to make a decision on the gear reduction mechanism in Q1. In Q4, B proposes to set the gear ratio so that a full rotation of the dial covers the whole measurement range. C performs the necessary calculations for that concept, and in Q8, asks others to consider the validity of his calculations, which leads B to think that they need 4 stages. In Q9, C brings up the issue of legibility of the dial, and asks others to interpret if the scale that would result from the gear ratio B is considering would be acceptable. A must have agreed with C's concern since she proposes a new dial concept—the dial rotating twice—in Q10. After the team considers that concept for about 45 seconds, C decides that 3 stages would be necessary if the dial rotates twice, and asks the others to make a judgment on her conclusion. B immediately agrees, and using 3 stages emerges as the decision. However, A is somewhat skeptical and challenges that decision in Q14 by trying to verify the assumption behind it. C addresses her concern, A does not object, so the consensus is reached and the decision is made. Q2, Q3, Q5, Q6, Q7, Q11, Q12, and Q13 play a role in the process by uncovering information and knowledge relevant to the formulation of Q4, Q8, Q9, Q10, and Q14 and D1.

This example implies a strong relationship, even a duality, between question asking and decision making. I will articulate it by proposing two axiomatic interdependencies:

1) Every question operates on decisions as premises since the questioner must make conscious distinctions regarding at least one or more of the following: the subject, object, and concept of the question. Questions are *formulated*. From the questioner's perspective, there is no such thing as an accidental question (even though questions might have accidental—unintentional and unanticipated—consequences, that is irrelevant to the formulation of the question and the questioner's motivation). In that sense, the questioner is bound to make decisions when formulating questions.
2) Conversely, every decision operates on questions as premises since decision making entails dealing with choices—decisions are devoid of meaning if a there is a single choice. Thus, there must exist a minimum of two choices, which constitute options that need to be contemplated, defined, compared and priced by the decision maker. I propose questioning as the enabling mechanism. Therefore, the decision maker is bound to question when making decisions.

From these interdependencies, it follows that the quality of the decisions a designer makes is coupled with the quality of the questions he/she asks, and that question asking and decision making should be given similar degree of consideration as topics of study under design cognition. Developing this approach might result in a new process unifying decision making and question asking, where decision making is viewed as taking place *during* question asking, and vice versa.

I believe that by constructing this argument I achieved my goal of conceptualizing a relationship between question asking and decision making that can serve as a vision for me while I study question asking in depth throughout the rest of this dissertation. I intend to advance that line of thinking and formulize that vision in the future by drawing on the findings of this work.

## 2.3 Review of Taxonomies of Questions

In this section, I explore existing knowledge on the nature of questions. I intend to apply that knowledge in laying out the foundations of a theoretical framework that would serve as an analysis scheme for the empirical part of this research, which entails observing designers while they design and analyzing their thinking. Taxonomies of questions are forms of knowledge on the nature of questions that are especially suitable for that role; categories of a taxonomy constitute natural units of a coding scheme that can be used in observation and analysis.

Therefore, in the next four sections, I review six relevant frameworks from five different disciplines: philosophy [Aristotle], education [Dillon 1984], artificial intelligence [Lehnert 1978], cognitive psychology [Graesser 1994], and design research [Kuffner 1990, Baya 1992].

### 2.3.1 From Aristotle to the Modern Scientist: Review and Classification of Research Questions

Dillon, an education researcher, reviews 12 categorization schemes for research questions published in the fields of education, philosophy, psychology and history [Dillon 1984], His goal is to understand more about the "kinds of questions that may be posed for research." He believes a comprehensive review might "serve to stimulate systematic work on the classification of questions for research in education and other enterprises of inquiry." He states that the utility of his approach can be viewed in three dimensions: understanding of inquiry, practice of inquiry and pedagogics of inquiry.

He argues that "understanding of inquiry" can take place at three different levels: the individual study, a corpus of studies and the enterprise of research in a given field. At the individual level, "a classification can reveal both the propositional and the contextual meaning of the question-answer pair represented by the study's problem and conclusion." Those meanings can be compared with the interpretations and applications of the findings. The results can be used to check for consistency between departure and arrival points, and to identify future research topics.

At the second level, he argues that "a categorical scheme of questions reveals what kinds of question are being asked, and thus what kinds of knowledge are yielded, about each of the various kinds of thing constituting the subject matter domain under review." In that sense, categorization of the questions used in a corpus of studies is a useful tool for organizing and communicating a "public conception" of a group of studies.

Similarly, at the third level, "classification can serve to understand the entire enterprise of inquiry in a given field." That understanding can aid researchers in gaining a more accurate conception of their research community. It would also be useful in outlining the types of questions that can be potentially posed, and the type of questions that are actually pursued.

The second dimension of Dillon's approach, "practice of inquiry," entails applying the understandings gained at the three levels outlined above to research practice; the design of the study is the focus as opposed to the understanding of the study. He argues that a hierarchical classification scheme would outline a procedure for the types of questions the researcher would be asking. Lower level questions would need to be answered before higher level ones. With the guidance of that procedure, the researcher would reach a position where he would know if a specific type of research question "can be safely asked" before posing it.

"Pedagogics of inquiry," the third dimensions of Dillon's approach, is the utilization of the understandings gained at the three levels in teaching. (Since Dillon is an education researcher, it would be plausible to assume that this dimension is the main motivation behind his study.) Different categorization schemes can be used to instruct students about the nature and function of research questions in order to expose them to the different principals associated with the schemes. That method can play a role in teaching the students how to construct their own research questions, and, thus, to frame their own research approach.

Dillon's review of the 12 categorization schemes yields mixed results. He finds that a significant portion of the taxonomies do not operate on specific and consistent differentiating principles. The principles used in forming the categories in most of the taxonomies are not made explicit by the authors, and examination of the taxonomies fails to reveal them. Therefore, Dillon argues that most of the published taxonomies have limited utility.

However, he perceives significant value in Aristotle's approach. As Dillon points out, Aristotle opens Book II of Posterior Analytics by proposing, "The kinds of question we ask are as many as the kinds of things which we know," and proceeds to identify four kinds of questions:

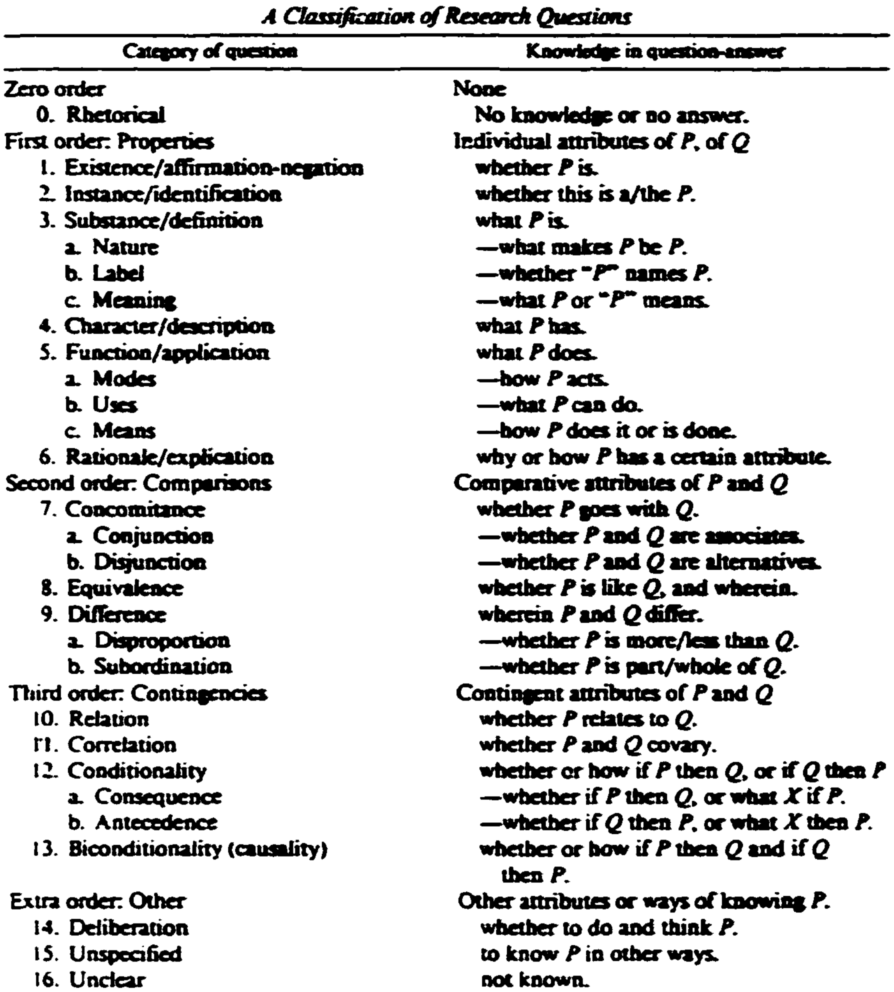
> "1) Whether the connexion of an atribute with a thing is a fact,
> 2) What is the reason of the connexion,
> 3) Whether a thing exists,
> 4) What is the nature of the thing."

As the four categories of questions illustrate, Aristotle's fundamental premise is to assume that our knowledge resides in the questions we can ask and the answers we can provide. After introducing the categories, he suggests a relationship between them by claiming, "When we have ascertained the thing's existence, we inquire as to its nature. When we know the fact, we ask the reason." Dillon interprets that relationship as a "sequence of inquiry", and illustrates the movement from question to question with the notation presented in Figure 2-6. Aristotle's words are on the right. Dillon's notation, suggesting categorial labels and a hierarchical movement between them, is on the left.

| | |
|---|---|
| **1. Existence/** | **When we have ascertained the thing's existence,** |
| **↓ Affirmation** | |
| **2. Essence/** | **we inquire as to its nature.** |
| **Definition** | |
| **3. Attribute/** | **When we know the fact** |
| **↓ Description** | |
| **4. Cause/** | **we ask the reason.** |
| **Explanation** | |

Figure 2-6. Dillon's interpretation of the squence of inquiry Aristotle argues for in Posterior Analytics. Aristotle's words are on the right. Dillon's notation, suggesting categorial labels and a hierarchical movement between them, is on the left.

Dillon then presents his own categorization scheme (Figure 2-7), which he states is being based on "Aristotle's few, short, and encompassing propositions." His scheme distinguishes between kinds of questions according to the knowledge about some phenomenon P entailed in the answer. It consists of three main orders that are representative of the sequence, or, rather, of the hierarchy, of questions proposed by Aristotle. That hierarchy is the basis of the questioning "procedure" Dillon suggests in the "practice of inquiry" dimension of his approach.



*A Classification of Research Questions*

| Category of question | Knowledge in question-answer |
|---|---|
| Zero order | None |
| 0. Rhetorical | No knowledge or no answer. |
| First order: Properties | Individual attributes of *P*, of *Q* |
| 1. Existence/affirmation-negation | whether *P* is. |
| 2. Instance/identification | whether this is a/the *P*. |
| 3. Substance/definition | what *P* is. |
| a. Nature | —what makes *P* be *P*. |
| b. Label | —whether "*P*" names *P*. |
| c. Meaning | —what *P* or "*P*" means. |
| 4. Character/description | what *P* has. |
| 5. Function/application | what *P* does. |
| a. Modes | —how *P* acts. |
| b. Uses | —what *P* can do. |
| c. Means | —how *P* does it or is done. |
| 6. Rationale/explication | why or how *P* has a certain attribute. |
| Second order: Comparisons | Comparative attributes of *P* and *Q* |
| 7. Concomitance | whether *P* goes with *Q*. |
| a. Conjunction | —whether *P* and *Q* are associates. |
| b. Disjunction | —whether *P* and *Q* are alternatives. |
| 8. Equivalence | whether *P* is like *Q*, and wherein. |
| 9. Difference | wherein *P* and *Q* differ. |
| a. Disproportion | —whether *P* is more/less than *Q*. |
| b. Subordination | —whether *P* is part/whole of *Q*. |
| Third order: Contingencies | Contingent attributes of *P* and *Q* |
| 10. Relation | whether *P* relates to *Q*. |
| 11. Correlation | whether *P* and *Q* covary. |
| 12. Conditionality | whether or how if *P* then *Q*, or if *Q* then *P* |
| a. Consequence | —whether if *P* then *Q*, or what *X* if *P*. |
| b. Antecedence | —whether if *Q* then *P*, or what *X* then *P*. |
| 13. Biconditionality (causality) | whether or how if *P* then *Q* and if *Q* then *P*. |
| Extra order: Other | Other attributes or ways of knowing *P*. |
| 14. Deliberation | whether to do and think *P*. |
| 15. Unspecified | to know *P* in other ways. |
| 16. Unclear | not known. |

Figure 2-7. Dillon's classification of research questions, distinguishing kinds of questions according to the knowledge about some phenomenon P entailed in the answer. Q stands for question.

The first order categories describe the properties of a phenomenon. The second order categories describe the comparative relationships, and the third order categories the contingent relationships between two phenomena. With regard to the relationship between the three orders, Dillon remarks that "the higher numbered categories and orders are classified as containing the lower by priority and increment of knowledge."

In order to determine the comprehensiveness of his classification scheme, he first demonstrates that all of the categories contained in the other schemes correlate with the categories contained in his scheme, and then extracts 924 "research questions" found in a sample of nine education journals for coding. He reports that his scheme accounts for

99% of the questions. He estimates the comprehensiveness of the other schemes by attributing the proportion of questions accounted by the corresponding categories of his own scheme[6]. Since none of the other schemes correlate with his scheme completely, that approach results in the comprehensiveness of the other schemes to be less than 99%. He reports Aristotle's scheme to be 89.1% comprehensive, and the other schemes to be 37%-83% comprehensive.

Based on these results, it can be said that Dillon's categorization scheme is one of the most comprehensive and representative frameworks for structuring our understanding of "research" questions.

### 2.3.2 AI Scientist's Approach: A Taxonomy of Questions for the purpose of Computer Simulation of Question Answering

Lehnert's work is aimed at laying out the theoretical foundations of a computational model—an artificial intelligence—that can answer questions [Lehnert 1978]. The computer program implementation of her model is called "QUALM." In her model, she treats question answering as a process that can be broken down into two parts: understanding the question, and finding an answer. The first part has to do with interpreting the question, the second with searching the memory of the artificial intelligence for the best answer. The first part of her approach requires the development of a taxonomy of questions[7]. Therefore, I will focus on and discuss the first part of her question asking process.

QUALM is based on Shank's theory of memory representation called "Conceptual Dependency" [Shank 1972]. In Lehnert's words:

> "Conceptual dependency is a representational system that encodes the meaning of sentences by decomposition into a small set of primitive actions. When sentences are identical in meaning, the Conceptual Dependency representations for those sentences are identical."

---

[6] Dillon argues that an indirect approach for determining the comprehensiveness of the other schemes is valid since he has proved his scheme to be encompassing of the other schemes as well as nearly all of the research questions in the data set, and that a scheme by scheme test is not necessary.

[7] Lehnert's taxonomy was not reviewed by Dillon.

Conceptual dependency assumes that "cognitive memory processes operate on the meaning of sentences, and not on the lexical meaning of those sentences." In other words, the fundamental operational mechanisms of memory are thought to be solely dependent on the conceptual meaning of what is being memorized, and to be independent of their lexical expression. For instance, the questions "Did Mary sell John a book?" and "Did John buy a book from Mary?" have similar conceptual representations.

As the above quotation points out, one of the basic structural elements of conceptual representations are "primitive actions." Conceptual dependency does not specify a finite set of primitives. However, the primitives it specifies are meant to constitute a small set so that its strength as a representation system is preserved. The following are some of the more important primitive actions Lehnert provides as examples:

ATRANS: The transfer of possession, ownership, or control.
PTRANS: The transfer of physical location.
PROPEL: The application of physical force.
MTRANS: The transfer of information.
MOVE: The movement of an animal involving a body part.
ATTEND: The act of focusing a sense organ toward some stimulus.
MBUILD: The thought process that constructs new information from old.

Another basic structural element of conceptual representations are "causal chains." They are used to establish causal relationships between the events described by primitive actions. For instance, when Mary falls and breaks her arm, gravity propelling Mary to the ground and Mary getting hurt constitute causally linked events, and the causal link is defined as "RESULT." Lehnert provides the following as being the six basic causal links in Conceptual Dependency:

RESULT: An event results in a state
REASON: Links mental events to nonmental actions.
INITIATE: A state or event initiates a though process (MBUILD).
ENABLE: A state enables an event.
LEADTO: Links two events such that the causal chain expansion is not explicit.
CANCAUSE: Modified LEADTO link where unspecified casual chain expansion is left out of the causal chain.

Based on the Conceptual Dependency framework, Lehnert breaks down the first part of her question asking process, understanding the question, into four specific stages: conceptual parse, memory internalization, conceptual categorization, and inferential analysis. The four stages of question comprehension are illustrated in Figure 2-8. What I would like to communicate with the diagram is that Lehnert treats question answering as a process, which can be implemented computationally.

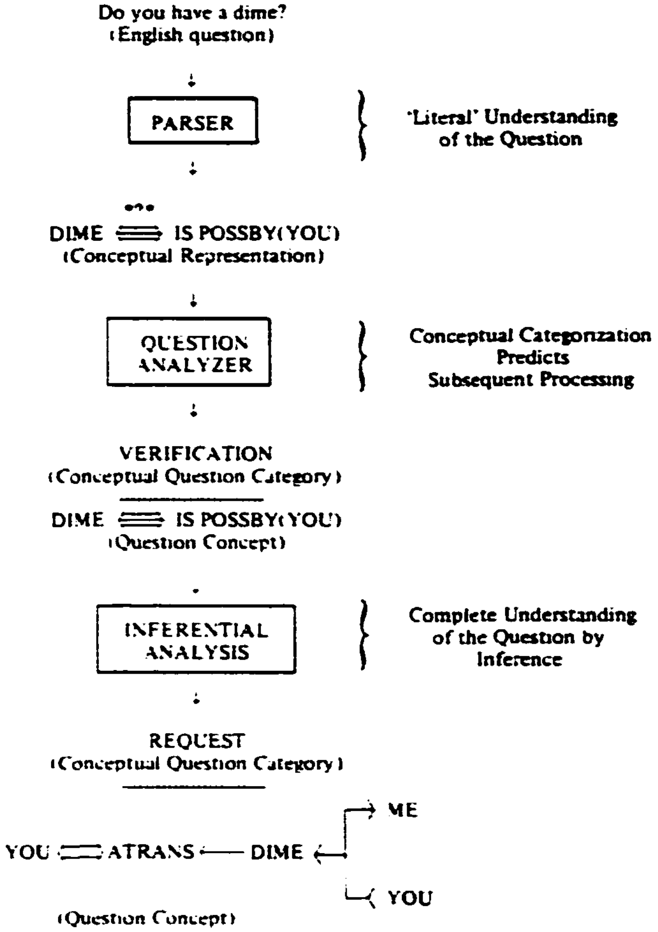


FIG. I.1. Stages of interpretation.

Figure 2-8. Based on the Conceptual Dependency framework. Lehnert breaks down the first part of her question asking process. understanding the question. into four specific stages: conceptual parse. memory internalization. conceptual categorization. and inferential analysis.

The input is the question, "Do you have a dime" expressed in English. The output is the understanding of the question with Conceptual Dependency principles; the question concept represented by the primitive action ATRANS, the transfer of possession, of the object DIME between the actors YOU and ME; and the categorization of the question as a REQUEST. In this example, there is no causal link in the question.

The parser is language dependent. It translates the question to an initial conceptual representation, which is internalized within memory by establishing the appropriate pointers to memory tokens for all references included in the question. The resulting conceptual representation can then be treated as being independent of a specific language.

The question analyzer "decomposes the initial representation into two descriptive components: a question concept and a conceptual question category." Question concepts are derived from the internalized question representations according to the rules developed for each category (for a detailed illustration of concept rules, see Lehnert 1972.) The conceptual question categories are determined by running the questions through a series of predetermined tests.

The resulting categorization of the question concept is tentative as it needs to be subjected to inferential analysis to ensure the question has been interpreted correctly. In the example being used, the question has indeed been incorrectly categorized as a verification question by the question analyzer. The inference analyzer corrects this mistake and categorizes it under the request category. The complete conceptual representation of the question, together with its question category, is illustrated at the bottom of Figure 2-8.

Lehnert's view is that the most important dimension of a question that needs to be interpreted for it to be understood and answered appropriately is its conceptual meaning. She also stresses that lexical categorizations differentiating between the so-called what, how and why questions "do not constitute a comprehensive system and are not motivated by anything greater than a desire to have a few general descriptive devices."

(The empirical part of this research independently arrives at evidence supporting her claim.)

Lehnert proposes 13 distinct conceptual question categories, and when viewed as a whole they constitute a taxonomy of questions articulating semantic differences. She thinks of the conceptual categories as "processing categories that are predicted by features of conceptual representation." When this statement is viewed in light of the fact that she is operating within the context of developing a question answering computer program, it would be accurate to qualify her categorization scheme as algorithmic and technical. The algorithm she implements in the question analyzer to determine the conceptual category of a question is illustrated in Figure 2-9.

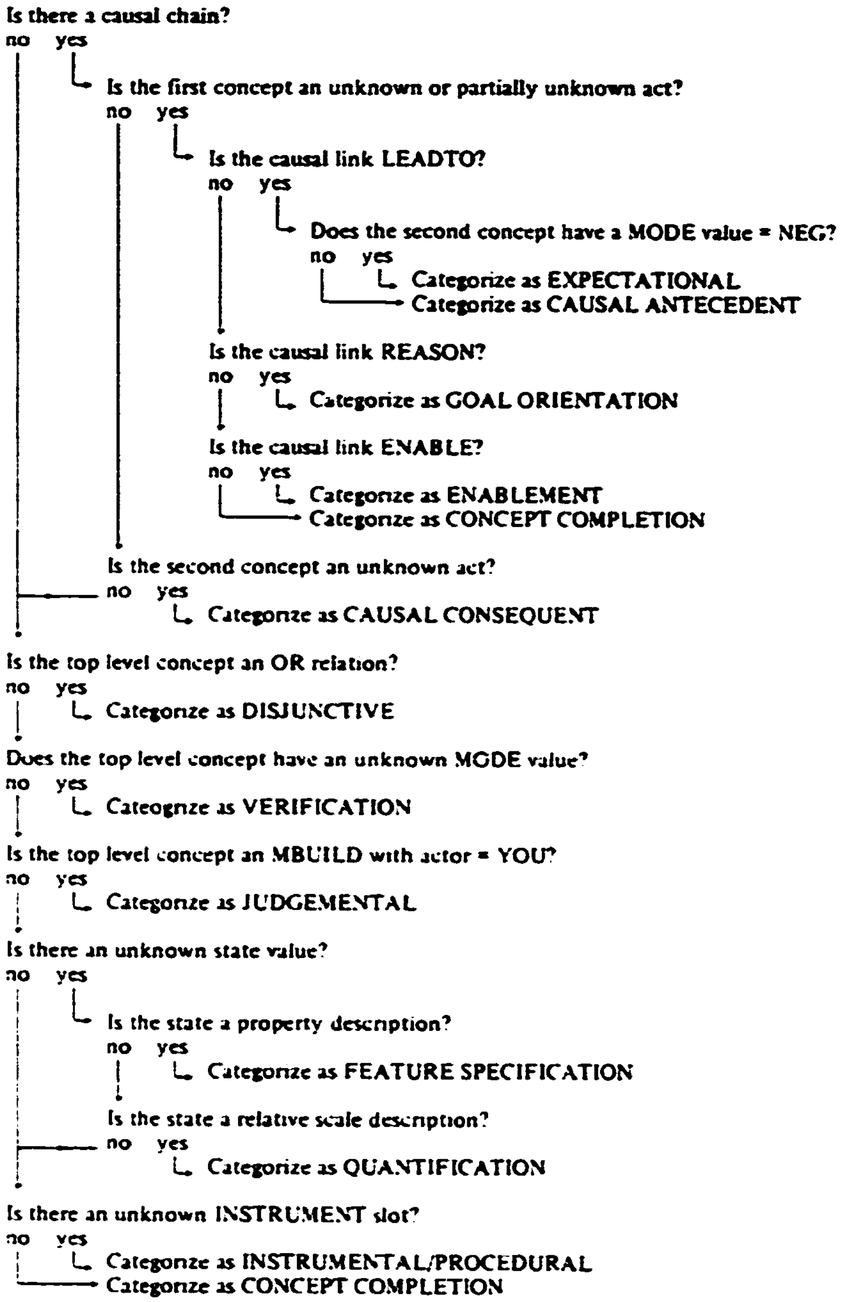


Figure 2-9. Lehnert's algorithm for determining the conceptual category of a question by the question analyzer.

The procedural test illustrated in Figure 2-9 results in a question being assigned to one of the following 13 conceptual categories (The description and examples are summarized from Lehnert's detailed discussion):

1. **Causal Antecedent**: The questioner wants to know the states or events that have in some way caused the concept in question. The causal link is LEADTO.
   Example: Why did the glass break?

2. **Goal Orientation**: The questioner wants to know the motives or goals behind an action (commonly referred to as the why-question). Goal orientation questions are a specific case of the causal antecedent questions in the sense that the reason behind the concept is mental. The causal link is REASON.
   Example: Why did John take the book?

3. **Enablement**: The questioner wants to know the act or the state that enabled the question concept. The causal link is ENABLE.
   Example: What did John need in order to leave?

4. **Causal Consequent**: The questioner wants to know the concept or causal chain the question concept caused. The causal link is LEADTO.
   Example: What happened after John left?

5. **Verification**: The questioner wants to know the truth of an event.
   Example: Did John leave?

6. **Disjunctive**: Verification question with multiple concepts.
   Example: Was John or Mary here?

7. **Instrumental/Procedural**: The questioner wants to know the partially or totally missing instrument in the question concept.
   Example: How did John go to New York?

8. **Concept Completion**: The questioner wants to know the missing component in a specified event (commonly referred to as the fill-in-the-blank question).
   Example: What did Mary eat?

9. **Expectational**: The questioner wants to know causal antecedent of an act that presumably did not occur (commonly referred to as the why-not question). The causal link is LEADTO.
   Example: Why didn't John go to New York?

10. **Judgmental**: The questioner wants to solicit a judgement from the answerer by requiring a projection of events rather than a strict recall of events.
    Example: What should John do to keep Mary from leaving?

11. **Quantification:** The questioner wants to know an amount.
    Example: How many people are here?

12. **Feature Specification**: The questioner wants to know some property of a given person or thing.
    Example: What breed of dog is Pluto?

13. **Request**: The questioner does not want to know anything, but wants a specific act to be performed.
    Example: Can you pass the salt?

### 2.3.3 Cognitive Psychologist's Approach: Considering the AI Taxonomy in the Context of Educational Goals

Graesser is interested in the cognitive aspects of question asking in an education context. His goal is to assess the influence of question asking on learning, and to identify mechanisms that generate questions [Graesser 1988, 1993, 1994]. He is also concerned with the role of questions in information systems [Graesser 1992].

He reports that even though education researchers and teachers seem to agree on the "virtues of being an inquisitive learner who actively exerts control over the materiel to be learned by asking questions," most students are not active but passive learners "who do not impose themselves on anyone with a question." Graesser points out that studies have shown that the questions students ask are "infrequent and unsophisticated," and "constitute approximately 1% of the questions in a classroom, at an average of one question per hour" [Dillon 1987, 1988; Flammer 1981; Kerry 1987]. The questions students ask tend to involve "the recall and interpretation of explicit material rather than questions that involve inferences, application, synthesis and evaluation." Also, attempts in facilitating the asking of more questions by the students have resulted in an increase in the number of unsophisticated questions. And finally, teachers do not fare much better in asking sophisticated questions; "less than 4% of the instructor generated questions are higher-level."

The taxonomy of questions Graesser presents is taken from Lehnert (see Section 2.3.2). Graesser adopts Lehnert's 13 semantic categories as they are, and adds five new ones. The categories he introduces are: "Comparison" (which he states was investigated by Laurer & Peacock, 1990), "Definition," "Example," "Interpretation" and "Assertion." Graesser does not provide a discussion on how the additional categories relate to the principles of Lehnert's taxonomy.

Graesser's main contribution is the application of the framework to empirical data [Graesser 1994]. He analyzes the frequency and the type of the questions asked by students during a series of tutoring sessions related to an undergraduate class on research methods. He focuses mainly on student questions, and not on tutor questions, as he claims they "reflect active learning."

He concludes that the frequency of the occurrence of a certain class of questions correlate positively with student learning ($R = 0.46$, $p < 0.05$ as measured by an examination score), and terms them "Deep Reasoning" questions, or "DRQs." DRQs consist of the following question categories: Instrumental/Procedural, Causal Antecedent, Causal Consequence, Goal Orientation, Enablement, and Expectational. He claims that, "such questions tap the steps and rationale in logical reasoning, in problem solving procedures, in plans, and in causal sequences."

In order to validate that claim and to generate a stronger argument for the correlation between DRQs and learning, Graesser considers DRQs in the context of Bloom's taxonomy of educational objectives in the cognitive domain. In Bloom's taxonomy, educational goals are organized into six hierarchical categories [Bloom 1956]. Accomplishing the higher level objectives requires the mastery of the lower ones. Graesser argues that deep reasoning questions are related to the higher level educational objectives, and therefore, are indicative of student learning.

He codes the student questions that were asked in the tutoring session according to Bloom's taxonomy, and tests for correlation between DRQs and the proportion of questions that are regarded as comparatively deep in Bloom's taxonomy (levels 2, 3, 4, 5, and 6). His analysis yields strong correlation (R 0.64, $p < 0.05$). He also reports that there is some correlation between the questions that are regarded as deep in Bloom's taxonomy and examination scores ($R = 0.35$, $p < 0.05$).

Graesser also reports on some other descriptive data that are relevant to the empirical dimension of this research. He reports that the students in the tutoring sessions generated 21.1 questions per hour, and the tutors generated 95.2 questions per hour (yielding a combined rate of 116.3 questions per hour for the student-tutor couple). This is very high compared to the 0.11-0.17 questions generated per hour in the classroom by each individual student (as reported by Dillon, Flammer, and Kerry). If only the DRQs are accounted for, the rates drop down to 4.6 questions per hour for students, and 15.2 for tutors (yielding a combined rate of 19.8 questions per hour). There are no data on the DRQ asking rates of students in classrooms.

### 2.3.4 Design Researcher's Approach: Two Taxonomies on the Information Needs and Handling of Designers

Even though the term "questioning" is often used in constructing and discussing design research paradigms, very few design researchers have directly studied the topic. Kuffner and Baya are one of the few researchers who have developed question-based research frameworks. Kuffner is interested in the information designers require to answer questions and to verify or refute conjectures about the design [Kuffner 1990, 1991]. Baya is interested in the nature of design information reuse and the role questions play in the information handling of designers [Baya 1992, 1996]. Their motivation is to aid the development of intelligent CAD tools.

Kuffner's framework is constructed specifically to draw out the relationship between questions and conjectures, and in a strict sense, does not constitute a taxonomy of questions. The main principle used for differentiating questions and conjectures is their verification attribute. If a conjecture is not followed with an immediate attempt at verification, it is called a "simple conjecture." If it is followed with an immediate attempt at verification, it is called a "conjecture with verification." Somewhat similarly, questions requiring only simple answers are called "verification questions," and questions requiring detailed answers are called "open questions."[8] Each question and conjecture is also categorized according to its "Topic," "Age of its topic," "Nature," "Confirmation" and "Validity." Topic is the "design object the questioner focuses on". Nature is dependent on the "type of information that the subject either seeks or presumes." Confirmation indicates if the question or conjecture is confirmed, and if so by whom or what. Validity "measures the accuracy of a conjecture."

Baya claims that "it is very natural for us to express our information needs in the form of questions," and he treats questions as identifiers of the content and the importance of the information designers seek. The question-centric framework he constructs reflects that thinking; the framework he uses to classify design information is identical with the framework he uses to classify questions.

Baya categorizes a question according to its "Descriptor," "Subject class," "Criticality," and "Level of detail." Descriptor refers to "the character or nature of the information being sought." It is almost identical to the "nature" class in Kuffner's scheme. Subject is "the subject of the sentence or the clause representing the questions." It is similar to the "topic" class in Kuffner's scheme. Criticality reflects the "measure of the impact asking of the question had on the overall goal of accomplishing a design." Level of detail is the level of detail of the information in the answer to the question.

Baya uses the taxonomy to analyze two design sessions where individual designers are asked to redesign a shock absorber. Due to the limited number of subjects, he treats the findings as descriptive data, which serve as a set of requirements for the development of DEDAL, a design information utility.

---

[8] It is relevant to note that this usage of the term "open question" is not consistent with its common usage by designers. I will address this issue in detail in the next chapter.

While commenting on the differences between his and Kuffner's frameworks, Baya makes three key observations:

- The size of the design problem will influence the range of questions one would encounter.
- Designers do not carry out design with a pre-determined set of questions. They raise questions as new information is needed.
- The questioning behavior is not random. New questions are being asked after reflecting on information received in answer to a question.

Even though these observations are information-centric—not all questions are asked to seek information—they are significant in the sense that they touch upon the notion of treating question asking as a process.

# 3 Development of a Taxonomy that is Comprehensive of the Questions asked while Designing

In Section 2.3, I took the first step in the development of a coding scheme that can be used to analyze questions by reviewing six taxonomies of questions.

In this chapter, I consider the comprehensiveness of those taxonomies when they are used to categorize questions asked while designing. My goals are to:

1) Discuss the appropriateness of treating the principles and question categories associated with the published taxonomies as analysis dimensions and units for studying the question asking behavior of designers.
2) Identify any dimensions of the question asking behavior of designers that are not addressed by those principles and categories.
3) If such gaps exist, propose new principles and categories that will address them.

Fulfilling these goals would constitute the second step in the development of a coding scheme, and result in a theoretical framework.

In Section 3.1, I provide the context for the observations that facilitated the consideration of the comprehensiveness of the published question categories, and the construction of new question categories. In Section 3.2, I address the issue of defining a question in a design context. In Section 3.3, I consider the comprehensiveness of the existing taxonomies in design situations, and identify a characteristic dimension of the question asking behavior of designers that existing taxonomies do not address. I then adopt one of the published taxonomies, and augment it by adding 5 new question categories in constructing a taxonomy of questions applicable to design situations. In Section 3.4, I consider four of the six taxonomies I reviewed and the one I developed together, and attempt to map them onto each other.

## 3.1 Context for the Observations on the Nature of Questions Asked While Designing

Before starting the discussion on the comprehensiveness of the published question categories, it is necessary to provide the context for the observations that formed the basis of my reflection and evaluation. As mentioned in Chapter 1, this dissertation has empirical and theoretical dimensions. A critical component of the theoretical dimension is the development of a taxonomy of questions asked while designing. The empirical dimension involves generating hypotheses from field observations, and designing and conducting experiments to test them. The connection between the two dimensions is the utilization of the taxonomy of questions as a coding scheme during the analysis of data collected from the experiments.

At a first glance, the theoretical and empirical dimensions might seem to be independent undertakings. However, my inquiry on the nature and categorization of questions, and the empirical research I conduct to test my ideas are corresponding endeavors. Even though one of my key start and end points is a theoretical framework, my process relies on establishing a dynamic dialogue between theory and empirical findings. The construction of a comprehensive and meaningful taxonomy is gradual and requires continuous reflection.

My process for maintaining that dynamic dialog is as follows: I start out with an existing taxonomy, which is the result of the contribution of researchers from different disciplines. I apply the taxonomy to the analysis of a design situation, and reflect on its appropriateness and utility in light of empirical data. The reflection allows me to make conceptual leaps in my understanding of questions. Each time I make a conceptual leap, I modify the taxonomy by refining existing categories and/or constructing new categories in order to incorporate the enhanced understanding. I then apply the modified taxonomy to another design situation to generate more empirical data and repeat the cycle. At the beginning of Chapter 4, I identify the three major steps that make up the empirical dimension of this research. Each step can be seen as one such cycle.

This cyclic approach produces a dilemma when it comes to presenting the findings that are embodied in the structure of the taxonomy: the gradual development of the understanding reflected in the taxonomy can be presented chronologically, or the final state of the taxonomy reflecting the most advanced understanding can be presented by itself. The first option is likely to be problematic and might possibly confuse the reader since Chapters 4, 5, and 6, which are centered around the three major empirical steps, contain crucial discussions on issues that are not related to the development of the question taxonomy, and are best communicated separately. Therefore, I choose the second option, and present the most advanced understanding on the question categories in Section 3.3.

The disadvantage of that approach is the absence of context for the discussion that I will present in this chapter. Naturally, the discussion will be much easier to interpret for the reader once he/she proceeds to read Chapters 4, 5, and 6. At this point, providing some background for the design situations I collected empirical data from might alleviate that limitation. I observed two types of design situations. The first one was a two week long real-life design project where a team of 4 graduate mechanical engineering students designed, prototyped, and raced a paper bicycle. The second one was a set of 90 minute long laboratory experiments where 14 teams of 3 graduate mechanical engineering students designed and prototyped a device that measures the length of body contours (the first 2 teams participated in the pilot version of the experiment.) The transcripts that I use to illustrate my arguments were created from the discourse of some of the teams who participated in the laboratory experiment.

## 3.2 Definition of a Question

Defining a question in a design context is challenging. Designers use a variety of communication mediums when engaged in design activity, and there are unique question posing opportunities associated with each medium. Gesturing [Tang 1991], interaction with hardware [Brereton 1999], sketching, and speech are some examples of potential communication mediums. Apart from such mediums, which require the active participation of an actor in the formulation of a question, elements of the design environment can constitute embedded question asking mechanisms. For instance, the mere presence of a person or an object in the environment could constitute a question (that, of course, would have to do with the person perceiving such elements in the environment as much as the presence of the elements in the environment themselves).

Researchers who conduct studies related to the role of questions in designing do not state explicit definitions for questions [McCracken 1990, Kuffner 1991, Baya 1992], nor do they explore the nature of questions at a comprehensive scale. They are focused in the pragmatic aspects of question asking. That is most likely because their primary interest is in understanding information flow and processing, and not directly in the broader cognitive aspects of question asking. (Those studies were discussed in detail in the previous chapter.)

There are, however, published definitions and deeper explorations of questions in other disciplines. In general, they refer to questions as inquiries that are expressed through written or verbal language. That understanding leads me to focus on the verbal exchanges that occur between designers. I omit the written exchanges since, in this study, I focus on observing and analyzing designing at the co-located team activity level, where written exchanges between designers are limited—if not nonexistent. Therefore, for the purposes of this study, I construct the following definition for a question:

> *In a design context, a question is a verbal utterance related to the design tasks at hand that demands an explicit verbal and/or nonverbal response.*

Even though this definition clearly limits the scope of my observations and their implications for reasons I mentioned earlier, I believe that it addresses one of the most common and influential modes of communication in group design activity, and, therefore, is a good starting point.

## 3.3 An Argument for the Search for the "Possible" and Its Characterization as Question Categories

The specific focus of this dissertation on question asking in design contexts allows me to identify an overlooked domain in the published taxonomies of questions. The common premise in their structure seems to be that a specific answer, or a specific set of answers, *exists* for a given question. Lehnert and Greaser also seem to assume that the answer is *known*—not necessarily by the person asking the question, in which case it would be a rhetorical question, but possibly by the person to whom the question is directed. Such questions are characteristic of *convergent* thinking, where the questioner is attempting to converge on "the facts." The answers to converging questions are expected to hold truth-value since the questioner expects the answering person to believe his/her answers to be true. Almost all of the categories of questions contained in Lehnert's taxonomy, including the ones Graesser refers to as Deep Reasoning Questions (DRQs), are converging in nature.

However, questions that are raised in design situations tend to operate under the diametrically opposite premise: that, for any given question, there exists, regardless of being true or false, multiple *alternative* known answers as well as multiple *unknown possible* answers. The questioner's intention is to disclose the alternative known answers, and to *generate* the unknown possible ones—regardless of their being true or false. Such questions are characteristic of *divergent* thinking, where the questioner is attempting to diverge away from the facts to the possibilities that can be generated from them. I find it useful to establish a terminology for those types of diverging questions, and find it appropriate to refer to them as "Generative Design Questions," or GDQs.

In light of the converging-diverging paradigm, a GDQ is similar to what is commonly referred to as an "open-ended" question by designers. Contrary to Kuffner's usage of the term, open-ended questions are generally regarded as having multiple answers, which satisfy the question in various degrees. Upon raising an open-ended, or a diverging question, the designer's role is precisely to tackle that quality of it by investigating and understanding how each answer satisfies the question, and by establishing criteria for favoring one answer over the others. That process of investigation, comparison and evaluation constitutes decision making in design. And, as I have argued for in the previous chapter, it does not necessarily take place after the question is posed; it also occurs while the questioner is formulating the question.

Therefore, a coding scheme for analyzing the questions asked while designing needs to account for the types of questions that fall under the GDQ concept as well if it is to be comprehensive. A good starting point is to adopt one of the more established taxonomies and augment it by adding GDQ categories. Two of the taxonomies reviewed in Chapter 2, Dillon's and Lehnert's, are especially insightful and comprehensive (since Graesser's taxonomy is an extension of Lehnert's, I will be referring to Lehnert only). Even though Dillon's taxonomy appears to be more structured, it is more appropriate for me to adopt Lehnert's for two reasons:

1) Lehnert's taxonomy has been proven to be applicable to coding questions in discourse, and its utility as a coding scheme has been enhanced by Graesser's discussion on DRQs.
2) Since Lehnert developed her taxonomy with the intention of creating an artificial intelligence that can answer questions, and actually implemented it as a computer program, it would be feasible to implement a framework that is based on hers as a computer program as well.

Therefore, I adopt Lehnert's taxonomy of questions and identify five GDQ categories as additions. The categories I propose are Proposal/Negotiation, Scenario Creation, Ideation, Method Generation, and Enablement. In the next section, will discuss and provide specific examples of each GDQ category. I will also illustrate the context in which each type of question occurs, and their significance, by providing transcripts from data segments extracted from some of the laboratory experiments.

### 3.3.1 Proposal/Negotiation

The questioner wants to suggest a concept, or to negotiate an existing or previously suggested concept. Even though those types of questions initially appear to fall under the "Judgmental" category, which covers questions where the questioner wants to solicit a judgement from the answerer by requiring a projection of events rather than a strict recall of events, upon further consideration, it becomes clear that a there is a fundamental conceptual difference between making a suggestion and soliciting a judgement.

An example of a Judgmental question is, "Do you think the wheel is more accurate?" The questioner is asking for the answerer's opinion on what should be done, and is not offering any opinion herself/himself. The answerer is expected to supply a single definitive opinion.

On the other hand, "How about attaching a wheel to the long LEGO piece?" is a Proposal/Negotiation question. The questioner is offering an opinion on a concept, and expecting the answerer to supply her/his own corresponding opinion(s), which would not be definitive. The questioner intends to establish a negotiation process by exchanging opinions, and to open up the possibility to new concepts. The suggestion of the new concept usually requires a consideration of the hypothetical possibilities the new concept can lead to.

Another example of a Proposal/Negotiation question is provided in Transcript 3-1, where Team 12 is considering a sensing concept for the measurement device. The consideration results in a new measurement concept.

| Time In | Time Out | Voice | Utterance |
|---|---|---|---|
| 0:23:49 | 0:23:50 | B | What do you call that? |
| 0:23:52 | 0:23:52 | C | Just a roller. |
| 0:23:52 | 0:23:52 | A | That would be a really interesting one. Just one piece you know the diameter of. |
| 0:23:54 | 0:23:54 | B | Roller... |
| 0:23:57 | 0:24:02 | C | It's basically a roller measurement. It's the same thing they use to lay out stuff on the streets |
| 0:24:05 | 0:24:05 | A | Or, you can make a... (cut off by C) |
| 0:24:07 | 0:24:10 | C | So basically do it in turns of fractions of circumference. |
| 0:24:11 | 0:24:16 | B | Okay, so we have a roller and then measure how many revolutions? |
| **0:24:17** | **0:24:25** | **A** | **Yeah, or you can have a series of Legos connected like a linkage that's really bendable, just kind of wrap it around like a tape measure, right?** |
| 0:24:26 | 0:24:27 | B | That's a good idea. It's another... |
| 0:24:28 | 0:24:41 | C | It's kind of an end-to-end thing you're talking about? So, you basically have two lengths that pivot, you know what I'm saying? So, you kind of flip one over the other and work your way around. |
| 0:24:45 | 0:24:45 | A | I was just thinking like...(cut off by C) |
| 0:24:47 | 0:24:45 | C | I was interpreting, trying to interpret what you're saying to mean something like this where you have something like this. |
| 0:24:56 | 0:24:56 | A | Oh, exactly. |
| 0:24:58 | 0:25:10 | C | That you could work your way around and flip one over the other so that you always have on length in contact with the surface that you're trying to measure. |

Transcript 3-1. Design team members A, B, and C are considering a sensing concept for a measurement device. The consideration results in a new measurement concept. The question highlighted in bold type is a Proposal/Negotiation question.

At the beginning of the transcript segment, C has already come up with the "roller" concept where the sensor is a wheel of known diameter that rotates freely on the surface to be measured. In the next 15 seconds, A and B converse with C, and learn how the roller works. When they understand that each revolution corresponds to a known distance, A transforms the concept to a linear domain and suggests the possibility of using a series of flexible linear linkages such a "bendable tape measure." A voices his suggestion in the form of the Proposal/Negotiation question highlighted in bold type in Transcript 3-1. C immediately responds to A's suggestion. He first makes sure he understood A's suggestion correctly, and then proceeds to refine the concept by negotiating its application method.

As can be seen in this interaction, Proposal/Negotiation questions are significant because proposing an idea in the form of a question promotes consideration and feedback, and negotiation promotes synthesis.

## 3.3.2 Scenario Creation

The questioner constructs a scenario involving the question concept and wants to investigate the possible outcomes. In a strict sense, such questions could be categorized under Lehnert's "Causal Consequence" category. However, Causal Consequence questions involve one causal chain of two conceptualizations—the second conceptualization is partially or completely unknown—joined by the LEADTO causal link. Scenario creation questions differ from causal consequence questions in two ways: there are multiple possible causal chains and linked concepts, and the causal link is CANCAUSE since the causal chains are hypothetical.

An example of a causal consequence question is "What happened when you pressed the pulley?" The questioner is assuming that when the person pressed the pulley, something specific happened. In other words, the person pressing the pulley led to a specific outcome, and the questioner wants to know what that was.

On the other hand, "What if the device was used on a child?" is a Scenario Creation question. The questioner wants to generate and account for as many possible outcomes as possible from the scenario(s) that can be constructed.

Another example of a Scenario Creation question is provided in Transcript 3-2, where Team 10 is evaluating a sensing concept for the measurement device. The evaluation results in the creation of a new measurement concept.

| Time In | Time Out | Voice | Utterance |
|---|---|---|---|
| 6:48:23 | 6:48:24 | A | We gotta keep this from rotating. |
| 6:48:30 | 6:48:32 | B | Can we like bend this? |
| 6:48:36 | 6:48:42 | A | Oh, what is this? Hey, check this out. I wonder if this has a rolling end? |
| 6:48:51 | 6:48:52 | A | Even works on clothing. |
| 6:48:53 | 6:48:55 | C | Yeah, it really's a matter of how tight you squeeze it. |
| 6:48:56 | 6:48:57 | B | We can do this. |
| 6:48:59 | 6:49:00 | A | That cantilever is wicked though. |
| **6:49:02** | **6:49:03** | C | **What about people who have hair?** |
| 6:49:04 | 6:49:05 | B | (laughing) Are you making fun of my hair? |
| 6:49:06 | 6:49:14 | C | (seriously) No. I'm saying that we have to measure...like this little wheel wouldn't work because it's not going to roll over long hair...even on my short hair it won't work. |
| 6:49:15 | 6:49:15 | A | Is it rolling? |
| 6:49:16 | 6:49:17 | B | No, a little bit. |
| 6:49:18 | 6:49:18 | A | Like, it slips. |
| 6:49:19 | 6:49:19 | B | You can't roll my...does it... (cut off by C) |
| 6:49:20 | 6:49:31 | C | Whereas the big one, or we could have an interchangeable roller, one that is pop-in for head, and pop-in for the hand. |
| 6:49:28 | 6:49:28 | B | Yeah. |

Transcript 3-2. Design team members A, B, and C are evaluating a sensing concept for a measurement device. The evaluation results in the creation of a new concept. The question highlighted in bold type is a Scenario Creation question.

At the beginning of the transcript segment, A, B and C are evaluating a sensing concept, where the sensor is a wheel of known diameter that rotates freely on the surface to be measured. A comments that the wheel even rolls on clothing. However, C realizes that it depends on how much pressure is applied on the axle of the wheel, and that it might slip. About 10 seconds later, C uses that insight to pose a Scenario Creation question, and wonders if the wheel would rotate without slipping on hair (the device will be used to measure the circumference of a human head). In essence, C constructs a new design requirement: the wheel should rotate freely and without slipping on hair. B then tests the device on his head, and reports that it indeed slips. At the end, C comes up with a new concept, which uses different size "interchangeable" wheels—the assumption being that a larger wheel would be less likely to slip.

As can be seen in this interaction, Scenario Creation questions are significant because accounting for possible outcomes generates and refines design requirements.

### 3.3.3 Ideation

The questioner wants to generate as many concepts as possible from an instrument without trying to achieve a specific goal. Such questions involve multiple possible conceptualizations and causal chains. The first conceptualization is partially unknown, and the second conceptualization is partially or completely unknown.

An example of an ideation question is, "Are magnets useful in anyway?" The questioner does not intend to achieve a specific goal by using the magnets. He/she does not have a purpose other than to generate as many ways of utilizing magnets as possible. The role of that question is illustrated in Transcript 3-3, where team 10 is considering magnets they came across while going through the hardware they were given to design and prototype the measurement device. The consideration results in a concept for holding the device while not in use.

| Time In | Time Out | Voice | Utterance |
|---|---|---|---|
| 6:29:34 | 6:29:35 | A | Wait, is this part of the kit? |
| 6:29:36 | 6:29:36 | B | Yes, magnets. |
| **6:29:37** | **6:29:40** | **A** | Hey there's magnets. **Are magnets useful in anyway?** |
| 6:29:43 | 6:29:44 | C | Yeah, if we wanna make an oscilloscope. (B laughs) |
| 6:29:48 | 6:30:07 | A | Let's try all the interesting pieces and see what we can do with them. Have an interesting piece section...I have no idea what it is...magnets...let's keep on moving them into big piles. |
| 6:30:10 | 6:30:12 | C | I don't even know why we have ball joints. |
| 6:30:23 | 6:30:34 | A | Let's see what they do here. They actually use these as rubber bands. That's kind of interesting...it would be cool to use our stuff. |
| 6:30:35 | 6:30:37 | C | I think these are just for these |
| 6:30:38 | 6:30:38 | B | What is that for? |
| 6:30:39 | 6:30:56 | A | Oh, that's interesting. Remember, esthetics count. Rubber band...(writing down the ideas)...uhm...squeeze handle, maybe we can do a squeeze handle. I don't know...Let's look through some of these cases. |
| 6:30:57 | 6:30:58 | B | There's something that bends. |
| 6:31:07 | 6:31:17 | A | Sockets just seem to stick out...Did you see the sockets do anything? They use sockets here to use the rubber bands to go on. |
| 6:31:19 | 6:31:19 | C | Oh. |
|  |  |  | (all three looking through the Lego manual) |
| 6:31:38 | 6:31:38 | A | Looks cool. |
| 6:31:41 | 6:31:41 | C | Let's make it (laughs). |
| 6:31:57 | 6:32:01 | A | Yeah, the magnet's sitting there, but it doesn't do anything. |
| 6:32:01 | 6:32:01 | C | They use magnets nere? |
| 6:32:03 | 6:32:10 | A | These are the magnets, right? With these tiny things clicked onto here. I'm not sure what they do. |
| 6:32:10 | 6:32:12 | B | I think it's just supposed to just hang stuff there. |
| 6:32:13 | 6:32:13 | C | So basically we have this thing, right? |
| 6:32:15 | 6:32:15 | B | Just hang stuff there. |
| 6:32:18 | 6:32:21 | C | That's his gun. He picks up at his pack and puts it... |
| **6:32:21** | **6:32:33** | **A** | So maybe we can use the magnet, maybe for as like a holder, so when you're done with it you just click it onto the wall or something...**What else can we do with magnets?** |

Transcript 3-3. Design team members A, B, and C is considering some magnets they came across while going through the hardware they were given to design and prototype a measurement device. The consideration results in a concept for holding the device while not in use. The question highlighted in bold type is an Ideation question.

At the beginning of the transcript segment, A identifies the magnets, and immediately poses an Ideation question in order to generate concepts for using them. It is important to note that at that point, A is acting without a specific goal; he does not have a specific role for magnets in mind. For a few seconds, they get distracted and focus on other "interesting" pieces like magnets, but they quickly come back to magnets and examine how they are used in the LEGO kit the parts came from. What they learn influences A to consider magnets as a part of a concept for holding the device while not in use. As soon as he generates that concept, he poses the same Ideation question in order to generate more concepts.

As can be seen in this interaction, Ideation questions are significant because operating without a specific goal frees associations and drives concept generation.

### 3.3.4 Method Generation

The questioner wants to generate as many ways as possible of achieving a specific goal. Even though such questions initially seem to be derivatives of Lehnert's "Procedural" category, they are fundamentally different. As Lehnert points out, "A Procedural questions asks about an act that was simultaneous with the main act of the question. If a question asks about an act that precedes the main act of the question, the question is either a Causal Antecedent or an Enablement question." A method generation questions falls into the second category since it asks about acts that precede the main act of the question. Then, according to Lehnert, it should be classified as a Causal Antecedent or an Enablement question. However, Causal Antecedent and Enablement questions involve a single causal link, whereas a method generation question has a completely known initial question concept and multiple possible and completely unknown secondary question concepts.

An example of a method generation question is, "How can we keep the wheel from slipping?" The questioner wants to generate secondary conceptualizations, which, if realized, will cause the initial conceptualization—keep the wheel from slipping. That question is clearly distinct from the causal consequence question, "What happened after you pressed the pulley?"

Another example of a Method Generation question is provided in Transcript 3-4, where Team 5 is generating methods for implementing an automatic readout of the measurement device. The evaluation results in the creation of several new readout methods.

| Time In | Time Out | Voice | Utterance |
|---|---|---|---|
| 6:05:01 | 6:05:09 | A | Let's brainstorm read-out methods. New topic. **However you measure it, how can you make it automatically readable?** |
| 6:05:16 | 6:05:17 | B | Okay, so have the audible clicking. |
| 6:05:19 | 6:05:21 | C | I think if we can do a visual. |
| 6:05:22 | 6:05:27 | A | Is there a rack and pinion? No, just simple gears. |
| 6:05:28 | 6:05:29 | C | We have some bevel gears though. I don't know if it's... |
| 6:05:32 | 6:05:44 | A | But if the spur gear rolls along a page, you can then whip out a tape measure and say, okay, this is how far it went, or something like that. You can make it like roll along something else. |
| 6:05:44 | 6:05:57 | B | That's why I was thinking if we wound up the string when you made the measurement then you just unroll the string and measure it...The rod I think is better. That's not elegant—unwinding some string and measuring it. |
| 6:06:08 | 6:06:13 | B | There might be way to make a magnet flip like 180 degrees every time. |

Transcript 3-4. Design team members A, B, and C are generating methods for implementing an automatic readout of a measurement device. The evaluation results in the creation of several new readout methods. The question highlighted in bold type is a Method Generation question.

At the beginning of the transcript segment, A invites the team to brainstorm readout methods. He immediately poses a Method Generation question, and sets their goal, which is to generate new methods for implementing an automatic readout, where the measurement the device takes is indicated in such a way that all the user needs to do is to look at the readout and read it off. The team responds, and within 60 seconds, generates 3 different methods.

As can be seen in this interaction, Method Generation questions are significant because operating with a specific goal generates a set of methods for implementing concepts.

### 3.3.5 Enablement

The questioner wants to construct acts, states, or resources that can enable the question concept. This category is the GDQ version of the original Enablement category Lehnert proposed, which Graesser labeled as a DRQ. What differentiates it from Lehnert's, and makes it a GDQ, is the questioner assuming the existence of multiple possible initial conceptualizations.

An example of a GDQ Enablement question is, "What allows you to measure distance?" when the questioner is indeed aiming at identifying resources for measuring distance. However, the same questions should be categorized as a DRQ enablement question when the questioner believes there is a single or a set of specific known resources of measuring distance. That differentiation can only be made by taking into account the context in which the question was raised. (The need to understand the context is true for categorizing any type of question, however, it is more pronounced in this specific case.)

Another example of an Enablement question is provided in Transcript 3-5, where Team 7 is generating resources that enable a measurement concept. The evaluation results in the identification of an existing resource and in the generation of a new one.

| Time In | Time Out | Voice | Utterance |
|---|---|---|---|
| **6:21:05** | **6:21:18** | **B** | So. **what goes around a circle and measures things?** You know...when you...like you ever...(pause)...Tape measure's pretty good. A tape measure! |
| 6:21:20 | 6:21:24 | C | I just keep thinking you just rotate this thing around. |
| 6:21:25 | 6:21:50 | A | Not necessarily. We can have something like let's say if we have a lot of little pieces joined like this. right...we can actually just put it around the hand. And it won't be...we'll have some minor error because it has spaces here. but if we do that we want each one like let's say this is one inch...these are all certain inches. certain lengths...we can just put that around the hand and measure how long it is. |
| 6:21:51 | 6:22:02 | B | I guess my comment. like things. my concern is that's a lot of parts. be we shouldn't really. we shouldn't really limit ourselves right now. But let's see what else we have. |

Transcript 3-5. Design team members A. B. and C are generating resources that enable a measurement concept. The evaluation results in the identification of an existing resource and in the generation of a new one.

At the beginning of the transcript segment, B poses an Enablement question in order to generate resources that can rotate and measure distance. It is important to note that he already has a measurement method in mind, rotation, and that he is looking for enabling resources. B immediately answers his own question by identifying a tape measure as a possible resource. Influenced from the tape measure, A then considers a different measurement method, conforming a series of linkages to the measurement surface, and generates a new resource that would enable it, straight LEGO pieces of known length connected at the ends. B briefly considers A's idea, and then returns to the Enablement question he asked in order to generate more resources.

As can be seen in this interaction, Enablement questions are significant because identification of multiple resources promotes surveying and learning from existing design features.

## 3.4 Comparison of Approaches

There are some striking similarities between the taxonomies I reviewed in Section 2.3. I already mentioned that Kuffner's and Baya's frameworks are rather similar. That is mainly because they both adopt highly focused and similar information-centric views. However, as Graesser argues while mapping Lehnert's taxonomy of questions to Bloom's taxonomy of educational goals, information seeking questions tend to have a lower significance in students' and, it can be argued, in designers' cognition. Therefore, understanding more about design cognition requires the construction of a taxonomy of questions that goes beyond accounting for information seeking questions. I have made an attempt in accomplishing that in the previous sections.

Therefore, at this point, it makes sense to look back and compare the classification schemes of Aristotle, Dillon, and Graesser, and the one I proposed. I already discussed how Dillon was inspired by "Aristotle's few, short, and encompassing propositions" when constructing his own scheme, and how the two map onto each other. I also discussed the origins of Lehnert's framework, and how it was fully adopted by Graesser and enlarged by the addition of five new categories. I remarked that Greaser's real contribution was to identify a class of questions as Deep Reasoning Questions, which are correlated with learning. Finally, I argued that in order for the modified taxonomy to be applicable to design situations, the addition of five more additional categories, representing divergent thinking, were necessary. I termed that class of questions Generative Design Questions. Thus, what we have so far is two parallel evolutionary threads on the taxonomy of questions. What remains to be done is to compare them and see if they map onto each other.

One way of conducting that comparison is to insert the five taxonomies into the columns of a table, and to attempt to align the rows—the categories—that are similar in nature. Mutually populated rows would point out synergy between the schemes. Table 3-1 illustrates the result of that comparison.

| ARISTOTLE | DILLON | LEHNERT | GRAESSER | ERIS |
|---|---|---|---|---|
| Existence (Affirmation) | Existence/affirmation | Verification | Verification | Verification |
| | Instance/identification | | | |
| Nature (Essence/Def.) | Substance/definition | | Definition | Definition |
| | | | Example | Example |
| Fact (Attribute/ Description) | Character/description | Feature Specification | Feature Specification | Feature Specification |
| | | Concept Completion | Concept Completion | Concept Completion |
| | | Quantification | Quantification | Quantification |
| | Function/application | Goal Orientation | Goal Orientation ■ | Rationale/Function ■ |
| | Rationale/explication | | | |
| | Concomitance | Disjunctive | Disjunctive | Disjunctive |
| | Equivalence | | Comparison | Comparison |
| | Difference | | | |
| Reason (Cause/ Explanation) | Relation | | Interpretation | Interpretation ■ |
| | Correlation | | | |
| | Conditionality & Causality | Causal Antecedent | Causal Antecedent ■ | Causal Antecedent ■ |
| | | Causal Consequent | Causal Consequent■ | Causal Consequent ■ |
| | | Expectational | Expectational ■ | Expectational ■ |
| | | Procedural | Procedural ■ | Procedural ■ |
| | | Enablement | Enablement ■ | Enablement ■ |
| | | | | Proposal/Negotiation● |
| | | | | Enablement ● |
| | | | | Method Generation ● |
| | | | | Scenario Creation ● |
| | | | | Ideation ● |
| | | Judgmental | Judgmental | Judgmental |
| | Rhetorical | | Assertion | |
| | | Request | Request/Directive | Request |
| | Deliberation | | | |
| | Unspecified | | | |
| | Unclear | | | |

Table 3-1. A visual comparison of the categories of five taxonomies of questions. Dillon's categories are an expansion of Aristotle's. Graesser's and Eris's categories are an extension of Lehnert's. ■ denotes the types of questions termed as "Deep Reasoning Questions" by Graesser. ● denotes the types of questions termed as "Generative Design Questions" by Eris.

The comparison results in seven distinct classes. The first four classes are the categories of Aristotle's classification scheme. The fifth class consists of the GDQ categories. The sixth class consists of the Judgmental category. The last class mainly addresses questions that do not truly seek answers, which do not constitute questions according to the working definition of a question used in this study, and unspecified questions that are not covered by any of the taxonomies.

It is logical to begin the analysis with treating Aristotle's categories as four baseline classes, and to see if the other three schemes based on Lehnert's work can be mapped onto them since Aristotle's scheme abides by a sound differentiating principle as well as a meaningful hierarchy. Lehnert's scheme abides by a sound differentiating principle as well, but lacks an order relating the categories. Dillon's scheme is based on Aristotle's, and, therefore, it does not need to be compared to Aristotle's to determine if it maps. Instead, his categories can be assumed to articulate and expand on Aristotle's broader categories, and constitute an extended baseline for comparison.

As Dillon points out, the differentiating principle between Aristotle's and his question categories is the extent of "knowledge about some phenomenon P entailed in answer." The hierarchy is the natural progression of that knowledge; lower category of questions contained in the initial classes have less knowledge in their answers than the higher categories of questions contained in the latter categories. The categories of questions contained in the last class have no, or unspecified, knowledge in their answers. Therefore, their positioning is irrelevant. Before discussing the appropriate positioning of the fifth class of questions, I will focus on the first four and the sixth classes and determine if the schemes map with respect to them.

Looking at Table 3-1, it is immediately apparent that Lehnert's scheme is missing the Instance category under the Existence class, the entire Nature class, the Equivalence and Difference categories under the Fact class, and the Relation and Correlation categories under the Reason class. On the other hand, Dillon's scheme does not articulate the Procedural/Instrumental, Enablement, and Judgmental categories that Lehnert's scheme contains. The rest of the categories in Dillon's and Lehnert's schemes map well.

The unaddressed Nature class in Lehnert's scheme is addressed in Graesser's by the Definition and Example categories, and the Equivalence and Difference categories under the Existence class by the slightly broader Comparison category. Even though Graesser's scheme does not directly address the Relation and Correlation categories, it can be argued that his Interpretation category partially maps onto them; interpretation questions can be thought to be exploring relationships and correlation between phenomena in order to construct causal explanations and projections. Also, the Enablement and Procedural/Instrumental categories not articulated by Dillon's scheme are most likely implied in Aristotle's Reason class, since such questions must assume and operate on the basis of causality.

Dillon's scheme does not address the Judgmental category proposed in Lehnert's scheme. That is mostly likely the result of Dillon's focus on research questions. When considered within the scope of Lehnert's framework, the Judgmental category is difficult to position among the other categories; all questions are judgmental questions to some extent since a question cannot be answered based purely on "fact" or with complete "objectivity." Therefore, I decided to treat the Judgmental category as a specific class, and to position it below the first five classes that are conceptually related.

In conclusion, at a fundamental conceptual level, the version of Lehnert's scheme Graesser augmented maps onto Dillon's, and thus, onto Aristotle's scheme. That is a positive finding as it indicates a strong degree of agreement in the thinking of the authors, and assures me that Lehnert's framework constitutes a sound basis for my analysis.

The fifth class of questions in Table 3-1 containing the Generative Design Questions is the contribution of this dissertation. It is not addressed by any of the other schemes. For the most part, that can be explained by the diverging-converging thinking paradigm I argued for in the previous section, where I made a fundamental distinction between questions that aim to converge on facts and questions that aim to diverge away from facts to the possibilities that can be generated from them. I believe that the classification schemes of Aristotle, Dillon, Lehnert and Graesser are concerned mainly with convergent questions.

One way of supporting that view, apart from interpreting the question categories directly, is to consider the motivations of the authors for constructing the taxonomies, and to assess if they aim to establish frameworks for understanding facts, or for creating possibilities from them. Aristotle's paradigm is Epistemological; as I remarked earlier, his main premise is: "The kinds of question we ask are as many as the kinds of things which we know." Thus, he focuses on what we know, on the existing, and not on the possible. Dillon explicitly states that his taxonomy is descriptive of "research" questions, and his interpretation of research activity seems to entail discovering and better understanding existing phenomena—paralleling Aristotle's paradigm.

And finally, Lehnert, strongly influenced by cognitive science, is ultimately interested in developing a question answering process, consisting of two separate processes for understanding questions and finding answers. The second process of "finding"—not creating—answers entails *retrieving* answers from existing memory structures. (Even though she mentions that multiple appropriate answers can be constructed for most questions using that procedure, that should not be taken to mean that possibilities can be created from known facts; it means that multiple known answers might exist and can be "found" in the memory structure.)

On the other hand, as I argued for in the previous section, the Generative Design Question categories I propose reflect divergent thinking. I therefore form a separate class of questions from them. However, it is not necessarily clear where that class should be positioned in Table 3-1, since its hierarchy is determined by the extent of knowledge in the answers. Does the knowledge in answers of GDQs encompass the knowledge in answers of the other class of questions? That is a problematic proposition since the purpose of GDQs is to *create* knowledge as opposed to discover or to construct it from fact, and it is inappropriate to guess at the extent of knowledge that is *yet to be* created before it is created. At this point, I can only hypothesize that GDQs, similar to DRQs, are correlated with learning, and also that both GDQs and DRQs are correlated with design cognition, and, thus, with design performance. Verifying that hypothesis would imply that the extent of knowledge in answers to GDQs is comparable to the extent of knowledge in answers to DRQs and to the types of questions in Aristotle's Reason class. I will address this hypothesis in the following chapters, where I will discuss the empirical dimension of this research.

# 4 Hypothesis Generation in the Field: Shadowing the Design Team

The empirical dimension of this research consists of three progressive steps:

1) Observation and analysis of a real-life design situation in the field for hypothesis generation.
2) Design of a laboratory experiment to test the hypotheses.
3) Evaluation and redesign of a pilot version of the experiment and the execution of the final version.

This approach to empirical design research—segmenting the research project into three distinct progressive steps—has been practiced in the Center for Design Research at Stanford University for over ten years, and has proven to be useful as it identifies a conceptual progression by providing definitions and outcomes for research steps.

Another approach I utilized—and, as I will discuss in the next chapter, augmented—is Tang's observational methodology for design research [Tang 1991]. The main principle of Tang's methodology is the iteration of a cycle consisting of the "Observe-Analyze-Intervene" phases, which advocates going beyond merely observing and describing design activity to constructing meaningful interventions to test the gained insights (Figure 4-1).

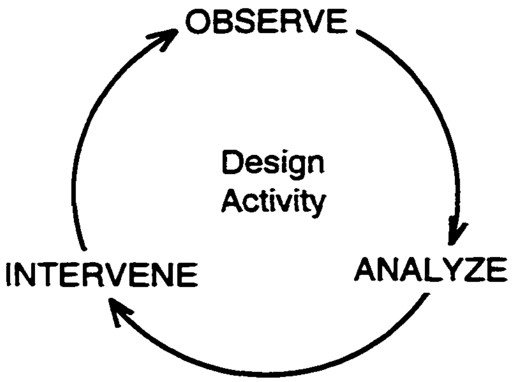

Iterative Approach to Empirical Design Research

Figure 4-1. Tang's observational methodology for design research [Tang 1991]. The main principle of Tang's methodology is the iteration of a cycle consisting of the "Observe-Analyze-Intervene" phases.

In order to use the two approaches in conjunction, I superimposed Tang's cycle on each of the three steps. While taking each step, I conducted multiple iterations of Tang's cycle. However, it is necessary to note that the nature of certain steps necessitates more emphasis on certain phases of Tang cycle than others [Figure 4-2].

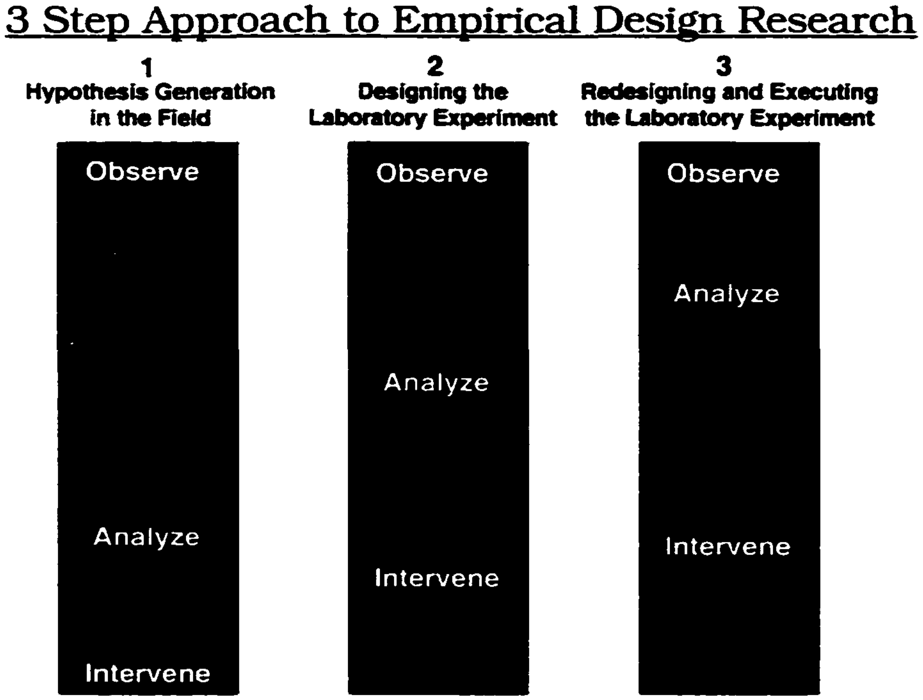


Figure 4-2. Tang's "Observe-Analyze-Intervene" cycle superimposed on the three steps of the empirical dimension of the dissertation. Each step entails multiple iterations of the cycle. Differences in the nature of the steps result in more emphasis on certain phases than other phases during each step. The relative sizes in the figure for each step are approximations for the time spent during each phase.

Specifically, during Hypothesis Generation, it is not useful, and may even be counterproductive, to focus on intervention. The main purpose is to observe and understand as much as possible about the design situation. When designing a Laboratory Experiment, the goal is to incorporate the understanding gained in Step 1 into experimental elements, observe their outcomes, analyze them, and reflect the findings in the design of a pilot experiment. The final step involves running the pilot experiment, observing and analyzing the outcomes, and redesigning it at least once so that the final version of the experiment accomplishes the intended intervention. Also in the final step, the final experiment is run and the data generated from it are analyzed in depth.

In this chapter, I discuss the first step of the empirical dimension of this research, hypothesis generation in the field.

## 4.1 Grounded Principle for Hypotheses Generation

In order to generate hypotheses in the field, I used a grounded approach, which involves identifying a real-life design situation, and documenting and capturing the activity in various formats for future analysis and synthesis. The main point is to ensure that the phenomena that will be identified for future study during field observations, and will be used in the construction of hypotheses are naturally occurring within design activity. While it is certainly useful for the researcher to bring his/her expertise, and thus biases, into the study, at the early stages of the research, it is imperative that he/she *perceives* rather than *proposes*. That approach ensures that the hypotheses are grounded in practice, and, therefore, relevant. If the researcher projects his/her expertise into the process early, the resulting hypotheses run the risk of being irrelevant to design activity. Naturally, verifying irrelevant hypothesis through experimentation accomplishes little in enhancing our understanding of and supporting design activity.

In other words, it is absolutely necessary to *study* the design activity first—regardless of how much insight one thinks he/she might have into what he/she is observing. Even though this principle is thought to be a basic understanding, it is very easy to drift away from it while observing "other" people designing, and to begin to develop a position on what "should be done." I believe there are two significant reasons for why that tends to happen:

1) Unlike social scientists studying social phenomena, design researchers studying design activity—a sociotechnical phenomena—tend not to be social scientists but designers, and, in many cases, engineers. And unlike social scientists, designers and engineers are *trained* to intervene and change systems as opposed to observe and understand them. That is not to say designers and engineers are not trained to observe and understand, but to say that their end goal, and hence priority and intent, is to intervene and change.

2) The nature of the activity under observation, designing, is simply engaging. If one were to observe swimmers swim, one would not necessarily be so tempted to start swimming himself/herself. However, if one is observing designers design, the feeling

is rather different as design activity has an encompassing human quality that invokes participation.

Therefore, applying grounded principles to empirical design research can require the researcher to constantly remind himself/herself of such implications while observing design situations.

## 4.2 Context of the Preliminary Observations

It is necessary to provide some background on the preliminary observations I conducted when generating hypotheses. Therefore, in this section, I will briefly discuss the setting for the observations, the designers I observed, and the design task they were engaged in.

### 4.2.1 The Setting: Mechanical Engineering 210, A Graduate Level Design Class

The setting for the preliminary observations was a graduate level engineering design class at Stanford University, Mechanical Engineering 210, Mechatronics Systems Design. The class lasts a complete academic year (3 academic quarters), and entails 30-40 students working in teams of 3-4 on industry sponsored design projects. Students are exposed to and master state of the art design processes and design support technology. In order to accelerate learning, a sociotechnical infrastructure, consisting of extensive coaching resources and collaborative design tools, is deployed. In order to facilitate the integration of resources, a "design loft" is used as communal space, where each team has a designated open work area.

During the first quarter of the class, students go through numerous warm-up "design exercises." At the end of the second month, they are introduced to a pool of industry sponsored projects, finalize their team formation efforts, and choose a project. Each industry sponsor provides conceptual, logistical and financial assistance through a project liaison and $15,000 budget to the team. At the end of 9 months, the teams are expected to deliver a functional prototype as well as detailed documentation of the design they have developed. The class has a history of producing highly successful projects (as measured by the success rate at the national Lincoln Arc Welding design competition all class projects are submitted to at the end of the year).

Apart from its educational value, this setting has also been acting as an observational platform and a test bed for researchers at the Center for Design Research. Since the class is structured to simulate real-life design environment—one that would be experienced in industry—the design activity that takes place in it can be treated as valuable data[9]. It can also serve as an experimental space where new design support tools developed by design researchers can be introduced and tested[10].

### 4.2.2 The People: A 4 Person Design Team

The ME 210 design team I observed consisted of 4 mechanical engineering graduate students[11]. Their backgrounds were also in mechanical engineering, and they were taking ME 210 as their core design class in the masters program. The team composition was in accordance with the design team-construction method developed by Wilde [Wilde 1997], which takes into account many psychological and academic descriptors of team members so that they have complementary modes of working. The team was unusual in one aspect: it consisted of three females and one male. The team members did not know each other before the class, and they formed the team using Wilde's team formation guidelines approximately two days before I began to observe them.

### 4.2.3 The Task: Design, Build and Race a Paper Bicycle

In ME 210, prior to the introduction of the industry sponsored projects, students go through a two week long introductory design exercise, which serves as a warm-up, and orients the students with the methods and technology that will be used during the rest of the class. For more than five years, the design task used in the introductory exercise had been to design, prototype and race a paper bicycle. The final prototype is expected to be built mainly out of paper components, meet some other constraints on weight, durability and stability. At the end of the two weeks, the teams enter a bicycle race with their

---

[9] Mabogunje discussed the validity of that claim in detail in his dissertation [Mabogunje 1997].

[10] There are ethical issues that require careful negotiation associated with any such effort.

[11] My observations, and the description of the course given in this chapter, are based on the version of ME210 offered in the 1998-1999 academic year.

prototypes, which takes place around a 400 feet circular track. Even though the duration of the exercise is somewhat short, I believe it is still a valid source of preliminary data for hypothesis generation.

## 4.3 Two Techniques for Capturing Design Activity in the Field and Generating Hypothesis

I relied on two techniques when gathering data in the field and generating hypotheses. Since both techniques are well established, I will not describe them in depth. Instead, I will make a few specific points regarding their use in empirical design research.

### 4.3.1 Ethnographic Approach: Shadowing the Design Team

Bucciarelli develops a social theory of design by using ethnographic techniques, and discusses their application to observing engineering design situations in his article, "An Ethnographic Perspective on Engineering Design," and his book Designing Engineers [Bucciarelli 1988, 1994]. As he points out, ethnographic techniques are an effective way of going beyond understanding designing simply by studying products to understanding designing by studying the design activity that creates them. In that regard, they can be considered as the mechanisms for abiding by the grounded principles I outlined in Section 4.1.

Before utilizing ethnographic techniques in the field as a design researcher, it is important to ensure that it is feasible to observe the design situation one wants to study. For instance, most commercial design projects operate under tight confidentiality regulations, and access to the "activity" will only be permitted in certain conditions. It is necessary to consider the potential effects those limitations might have on the outcome of the study as some situations might simply not permit the level of access necessary to generate significant insights. However, it is important to keep in mind that in most cases, such limitations can be negotiated and reduced over time.

Fortunately, the setting for the field observations in this research, ME 210, did not pose any significant limitations of that nature, as graduate students tend to be open to being observed. However, even though the class strives to "simulate" real-life design situations, the design activity the students engage in still tends to possess an academic quality. It is possible to view that as a tradeoff between access and reality. As compared to a commercial setting, in the academic setting, the researcher has nearly unlimited access, but less real-life data. However, I believe, and as Mabogunje points out, that does not negate the validity of the observations carried out in academic settings such as ME 210.

Therefore, for this study I, together with a colleague, "shadowed" a four person ME 210 team during the paper bicycle project. Upon spending a brief amount of time with each team in the class, just when the project was beginning, we choose the one we thought would be the most accessible. The team agreed to inform us in advance of the time and place of every informal or formal group meeting—design sessions—that involved at least three of the team members. Over the two week duration that the projected lasted, we were notified of over nine such design sessions, and observed all of them by using basic ethnographic techniques.

### 4.3.2 Video Interaction Analysis: Generating the Hypotheses

Another technique we employed in conjunction with ethnography was to capture the interaction during the design sessions with a video recorder. Fundamentals of video interaction analysis and its use in design research have been discussed in detail by Tang and Cross [Tang 1991, Cross 1996].

A significant difference between the two methods is that, as an ethnographer, the researcher relies on his own senses and strives to document as much of his perceptions as possible through note taking during and after the observations, whereas when using the video camera, the researcher relies on the audio and video information the video camera can capture. Therefore, each method can be thought to document the activity through a different "lens." That is desirable since, if used in conjunction, the data generated by each technique can be complementary—the findings generated with one method can add meaning to the findings generated with the other.

Another significant difference between the two techniques is that the information captured with a video recorder can be replayed. This has two implications; video data can be shared and independently analyzed by other researchers who did not directly observe the captured design activity; and when aiming to generate hypothesis, video data can be jointly analyzed and used as a means to facilitate unstructured reflection.

The first implication widens the scope of the analysis that can be done with the data. As was the case with the data used in producing Analyzing Design Activity [Cross 1996], the videotapes can be sent to groups of researchers, analyzed and interpreted by them, and their findings can be compared and synthesized into a collective understanding.

In phrasing the second implication, what I mean by unstructured reflection is a form of brainstorming, where several researchers watch the videotapes together, and, *while* doing so, speculate freely on any aspect of the activity that might attract their attention with the intent of generating hypotheses. This widens the possibilities in generating hypotheses as the interaction between researchers is very likely to stimulate their ideation process. That is how the video data collected during the paper bicycle project were utilized in this study.

## 4.4 Findings of the Field Research

The findings of the initial step of the empirical dimension of this research, observation and analysis of a real-life design situation in the field, lend themselves to discussion in four sections. In the first section, I evaluate the effectiveness of the two techniques of observation and analysis addressed in Section 4.3. In the second section, I focus on the outcomes of the observation and analysis, and highlight four key observations. In the third section, I derive three testable hypotheses by considering the key observations together with the conceptual framework I developed on the nature of questions in Chapter 3. And finally, in the forth section, I synthesize the phenomena outlined in the hypotheses into an analytical framework for understanding and measuring design performance.

### 4.4.1 On Capturing Design Activity in the Field

The two techniques I discussed in Section 4.3, when used in conjunction, proved to be highly effective in capturing design activity in the field. Even though I cannot comment on their individual effectiveness, I can say that using them in conjunction enhanced the accuracy and depth of my observations by providing me with different levels of granularity and focus. I will illustrate that point by characterizing two common scenarios that arise when analyzing data.

There were "tacit" elements of the interaction that were not necessarily reflected in the videotapes, but were visible if one were observing the interaction in person. For instance, it was possible to gain an idea of the general "mood" of the team by watching the videotape of a meeting. However, it was difficult to identify how that mood had developed to its recognizable state. On the other hand, witnessing the interaction in person enabled me to sense and understand more about the sentiments of the individual team members, and how those sentiments led to a collective mood. What I refer to as the mood of the team reflects strongly to the team's actions—its motivations, questions and choices—and, therefore, is highly relevant and needs to be observed.

Another element of a tacit element of the interaction was what took place *outside* of what the camera was capturing. The background environment and activity influenced what the team was doing. Also, there were stretches of time where one or more team members moved away from the others, and could not be captured with the video camera. What they were doing while they were away from the others, and the significance of those actions could only be interpreted by being there.

Conversely, observing interactions that were subtle, or simultaneously happening with other interactions, in person proved to be difficult since, as an ethnographer, it was only possible to focus and observe a limited number of actions at any given time. However, the video camera does not have the same limitation as an instrument; every visual interaction within its plane of focus is recorded at the same resolution, and the interaction that has been recorded can be replayed for an unlimited number of times.

Therefore, while analyzing videotapes, I was able to notice interactions which I had not noticed when observing in person. For instance, it was possible to miss what a team member was doing with the prototype from a previous project while trying to follow what another one was sketching on the board. It was only when I viewed the videotape later that I noticed the interaction between the team member and the prototype. Also, in many instances when team members were talking simultaneously within the team, or having separate one on one discussions, it was impossible to follow all of what was being said. Analyzing such situations from videotape enabled me to identify significant ideas, questions, and decisions that were discussed which I had missed as an ethnographer.

What I have reported above indicates that design activity is inherently *rich* and can be observed and characterized at various levels. The spectrum of activity and environment depicted in Figures 4-3, 4-4, and 4-5 reflect only a fraction of that richness. The figures contain frozen "frames" from sections of the video data corresponding to progressive phases of the paper bicycle design project that I observed.

Figure 4-3. Frames from video data: The paper bicycle design team conceptualizing in their team space ( on the left) and the class design space (on the right).

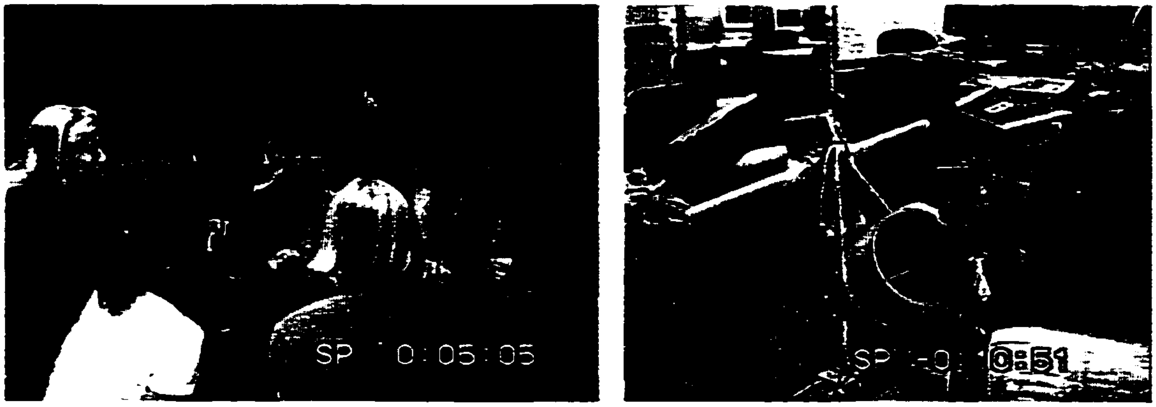


Figure 4-4. Frames from video data: The paper bicycle design team exchanging ideas and best practices with another team (on the left). The paper bicycle design team prototyping their design (on the right).

Figure 4-5. Frames from video data: The paper bicycle design in the final design review with class TA's and instructor (on the left). The final paper bicycle prototype of the design team (on the right).

The point I want to make with the figures is that by "shadowing" the design team in the field for the duration of the design project, utilizing both ethnographic and audiovisual recording techniques, and analyzing my field notes in conjunction with the videotapes, I believe that I have been able to approach dealing with the "totality" of the design activity and to gain a fundamental understanding of what took place.

### 4.4.2 Key Observations

When analyzing my notes and the videotapes, I focused solely on the role of questions in the interactions between the design team members. What I observed in the team's question asking process lie at the very heart of the arguments I present in this dissertation.

I made four significant observations:

**O1:** The design team members spent a significant portion of their time asking and discussing questions related to the design tasks at hand. They used questions in order to: mediate their social interaction, verify and clarify facts and each others views, seek new information, reason about and explain phenomena, and generate ideas. (This observation alone convinced me that question asking was a worthy subject to study.)

**O2:** Meetings during which the team seemed to ask more "good" questions (at that point the definition of a good question was highly intuitive and subjective to me) yielded more progress in terms of the insights the team seemed to gain and the discoveries they made.

**O3:** Working with existing artifacts and prototyping hardware seemed to have an effect on the types of questions the team members asked. Initially, when hardware was not present or rarely referenced, their questions were more conceptual and abstract, and required long answers and lead to detailed discussions. Toward the end, when they were discussing existing artifacts and working with prototyping hardware, their questions were much more specific and focused. (I was able to witness this trend since we had videotaped all of the meetings for the complete duration of the project.)

**O4:** However, distinguishing questions in discourse was difficult. I repeatedly found myself rewinding the tape after viewing the activity that followed a question just to make sure what I initially thought was a question was indeed a question.

### 4.4.3 Three Testable Hypothesis

I used the key observations outlined above, together with the conceptual understanding I gained while developing a taxonomy of questions applicable to design activity, as a basis for generating testable hypothesis.

A good starting point was to determine elements of question asking—apart from the questions themselves—that could be characterized and formulized. I postulated two such elements: the nature of a question, and the timing of a question. Considering those conjectures in light of the first observation, I wondered if they could be treated as descriptive characteristic of the design process. In other words, can a person who is exposed to those two characteristic elements, together with the questions, reconstruct the fundamentals of how the team structured its design tasks? This constitutes my first hypothesis.

Considering the second observation, my curiosity shifted to possible relationships between the frequency of questions and performance. Do designers who ask more questions simply perform better? And if so, can questioning be treated as a real time design team performance metric? This constitutes my second hypothesis. I believe this one is of particular importance since researchers in the field have, as of yet, not been able to identify real-time performance metrics even though there is a conviction that they are badly needed. There are various performance metrics which evaluate products of the activity such as sketches, documentation [Mabogunje 1997], and designed artifacts, however, when compared to a real-time metric, they are of lesser utility in terms of understanding and managing an ongoing design project.

The third observation led me to consider the potential effects of working with prototyping hardware on the question asking behavior of designers. I assumed that the observed changes in the nature of the questions asked would be reflected in their "type" if they were to be categorized according to the framework developed in Chapter 3. By integrating that assumption with the third observation, I postulated that the types of questions design teams ask change when they transition from working in the absence of hardware to working with hardware. This constitutes my third hypothesis.

To summarize, O1, O2 and O3 led to the following testable hypotheses:

**H1:** Question timing and type are descriptive characteristics of design cognition and process. When the set of questions a design team asks during a design project is considered as a whole, the timing and nature of those questions point at the fundamentals of the knowledge and rationale the team uses for breaking down and structuring the project into design phases. Question timing and type are informative enough to serve as a roadmap to the design thinking and process of the team.

**H2:** Overall question asking rate is related to design team performance and can be taken as a design performance metric. There is a strong correlation between the frequency of questions and design team performance.

**H3:** Question asking behavior of design teams is influenced by their access to hardware. The types of questions design teams ask change when they transition from working in the absence of hardware to working with hardware.

### 4.4.4 A Framework for Measuring Design Performance

When viewed together, the phenomena outlined in the above hypotheses form the hierarchical elements of an analytical framework for understanding and measuring design performance (Figure 4-6). Each phenomenon can be viewed as a descriptor of a higher encompassing phenomenon. The feasibility and accuracy of any descriptor as a performance metric increases with decreasing level, since lower level descriptors possess more detail, and are easier to identify, and, hence, measure.

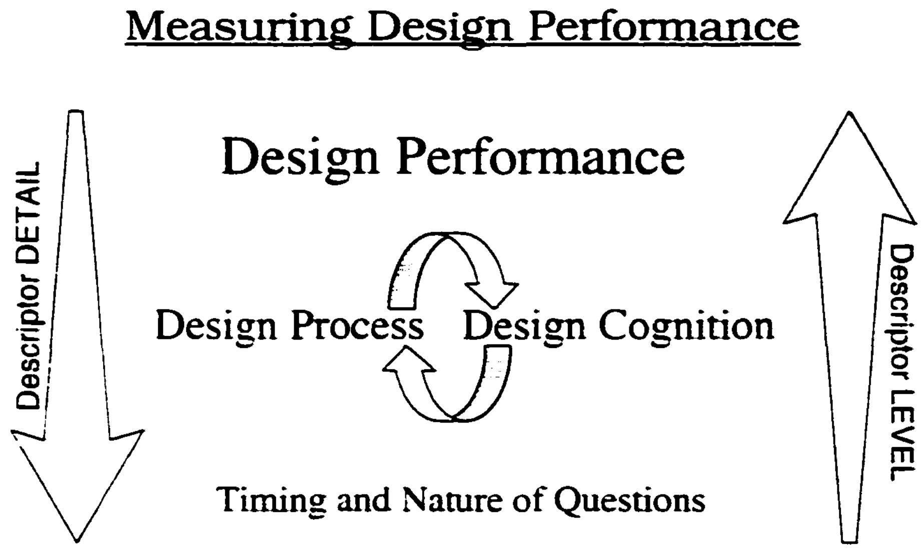


Figure 4-6. When viewed together, the phenomena outlined in the hypotheses form the hierarchical elements of a framework for understanding and measuring design performance. Validation of the hypotheses would imply the validation of this framework.

It is important to note that I consider design process and design cognition to be descriptors of the same level. They are strongly dependent on each other in the sense that one directly feeds the other in a cyclic fashion. Individual designers, and, as I have argued for in an earlier article [Eris 2002], product development organizations, extract and construct new design processes from existing design knowledge and thinking, and the resulting design processes form the basis for new design knowledge and thinking.

The implication is that, in the context of measuring design performance, observing and testing the relationship between one of them and question asking can be considered to be sufficient in generating indirect evidence for the relationship between the other and question asking. However, in general, design processes of teams and organizations are much more visible, and, thus, easier to observe and track than their design cognition. Therefore, when dealing with H3 in the following stages of this research, I focus on and observe only the design processes of the teams when analyzing data.

Finally, since the elements of the framework I propose for understanding and measuring design performance are hierarchical, validation of the hypotheses would imply validation of the framework as well.

# 5 Designing the Intervention: Differentiating Designing from Problem Solving

The second step of the empirical dimension of this research is the design of a laboratory experiment to test the hypothesis generated during the analysis of field observations. In the initial section of this chapter, I identify seven design requirements under three experimental design criteria that need to be met for the experiment to test the hypotheses. The hypotheses outlined in Chapter 4, the conceptual framework of questions presented in Chapter 3, and experimental considerations specific to design research discussed in this chapter all serve as natural design criteria. In the second section, I discuss each requirement in depth, and propose ways of meeting them. In the final section, I specify a design exercise that meets the requirements.

## 5.1 Deriving Requirements for the Design Experiment

The seven requirements under the three experimental design criteria are the following:

### Taxonomy Related Requirements

**R1:** The design experiment should promote realistic question asking processes from teams so that the application of the taxonomy of questions, which itself is based on data from realistic question asking processes, is meaningful.

### Hypotheses Related Requirements

**R2:** Definitions and metrics for the phenomena outlined in the hypotheses should be developed prior to the execution of the design experiment.

**R3:** The design experiment should incorporate an intervention that promotes a clear distinction between designers working with and without hardware.

Design Research Experimentation Related Requirements

**R4:** The design experiment should promote designing as opposed to problem solving.

**R5:** The setting and scenario of the design experiment should allow for the insertion of control elements associated with the hypotheses without overconstraining the activity (quasi-control as opposed to tight control).

**R6:** The design experiment should facilitate the testing of all hypotheses in a single experiment.

**R7:** The data collection methods used in the design experiment should result in data that can be analyzed qualitatively as well as quantitatively.

In each of the following three subsections, I will focus on a criteria and present the rationale behind the requirements that fall under it.

### 5.1.1 Taxonomy Related Requirement

**R1:** The design experiment should promote realistic question asking processes from teams so that the application of the taxonomy of questions, which itself is based on data from realistic question asking processes, is meaningful.

R1 reflects the understanding I gained while developing the taxonomy of questions. If the question asking processes of the teams in the experiment are indeed realistic, and if H1 is true, then it should be possible to view and explain the questions generated by the teams in terms of the categories of the taxonomy; the distinctions embodied in the taxonomy should serve as a comprehensive coding scheme for analyzing data.

In other words, if the taxonomy I develop is indeed comprehensive and rich, when applied to a design situation simulating realistic design activity, each of its categories serving as analysis codes should receive multiple hits. And conversely, if the situation I construct for the experiment indeed simulates realistic design activity, when coded by the categories of a comprehensive and rich taxonomy, it should incur multiple hits on each category. However, the coding scheme resulting in multiple hits per category alone does not guarantee either that the design situation simulates realistic design activity or that the taxonomy is rich and comprehensive. That can only be ensured through qualitative assessment.

### 5.1.2 Hypotheses Related Requirements

**R2:** Definitions and metrics for the phenomena outlined in the hypotheses should be developed prior to the execution of the design experiment.

R2 necessitates the development of working definitions and metrics for the phenomena outlined in H1, H2 and H3 prior to conducting the experiment. Since the phenomena constitute the units for analysis, it is important that they are characterized clearly so that the analytical framework for understanding the data is soundly in place before data collection takes place. The phenomena under investigation are:

1) Question Timing and Frequency
2) Question Type
3) Design Phase
4) Design Team Performance

**R3:** The design experiment should incorporate an intervention that promotes a clear distinction between designers working with and without hardware.

R3 aims to ensure the testing of H3 by requiring experimental control elements that result in a distinction between design teams working with and without hardware. The rationale behind R3 is to recreate, analyze, and thus, better understand the observed relationship between the question asking behavior of the paper bicycle design team and its use of hardware.

At the beginning of the project, the team did not bring any prototyping hardware to their meetings, and rarely felt the need to reference or examine existing paper bicycles. (There were several paper bicycles on display in the design loft in which the team held most of their meetings that were built during the previous offerings of the class). During those meetings, the team operated predominantly at a conceptual level. Approximately halfway through the project, they started building physical prototypes. Shortly after, especially when they ran into problems and were stuck, they began to pay close attention to the bicycles from previous years, examined their design principles, and borrowed and incorporated any ideas they deemed useful.

As outlined in O3, initially, when hardware was not present, the questions the team asked were more conceptual and abstract, and required long answers and led to detailed discussions. When they started working with prototyping hardware and interacting with the existing artifacts, their questions became much more specific and focused.

There probably are various causes for the shift in the question asking behavior of the team other than the team's interaction with hardware. For instance, it is possible that the shift might be a temporal phenomenon related to the life-cycle of a design project. Regardless, H3 focuses on the influence of the access to hardware, and R3 requires the insertion of control elements that recreate the type of interaction the paper bicycle team had with hardware in the experiment.

### 5.1.3 Design Research Experimentation Related Requirements

R4 through R7 are methodological requirements specific to design research experimentation. They need to be met for the experiment to be considered a contribution to design research. In formulating them, I take the position that the main prerequisite of a design experiment—independent of the hypotheses it is attempting to validate—is to convincingly simulate a real-life design situation.

**R4:** The design experiment should promote designing as opposed to problem solving.

In formulating R4, I make a distinction between problem solving and designing, and advocate that the experiment should promote the latter. Designing and problem solving are often treated as synonymous parametric processes. There is a prominent belief in society that engineers "solve problems" when they design. My view is that even though there is truth in such beliefs, designing and problem solving are fundamentally different in nature. One can choose to look at the world—let alone engineering—through a lens which influences one to perceive most things as problems that need to be solved, and that can be very useful. However, if that paradigm is pushed too far, it can be rather limiting because there are many situations in life, and in engineering, which require a more open-ended consideration. I believe the term "designing" addresses that very issue by constituting a meta-paradigm that encompasses problem solving, and relates it with other concepts such as perception and communication.

More specifically, in engineering design theories, it is common to assume that designing (and problem solving), transpire in two distinct domains; the domain of requirements and the domain of solutions (also referred to as the so-called requirement and solution spaces). It is also common to assume that constructing relationships between the elements contained within the two domains constitute the design (and problem solving) activity.

Even though I have reservations about subscribing to such an approach, which assumes the existence of the requirement and solutions domains, I will utilize it to make my point. Building on existing views regarding the negotiated nature of design requirements [Buccarelli 1994, Minnemen 1991, Eodice 2001], I argue that, in a problem solving context, requirements are given and are treated as such by the problem solver, whereas in a design context, they are negotiated, and even constructed, by the designer. I also argue that, in a problem solving context, solutions are final and take on a static role once formulized, whereas in a design context—borrowing from existentialist thinking—they are never reached, and even, never truly *exist*.

A simple example is to consider if the activity an engineering student who is engaged in solving a problem in a statics course—no matter how advanced the course might be—and the activity a practicing design engineer in industry who is designing a crane is engaged in are conceptually the same. I believe that even though the two activities have similarities, they are not conceptually the same. It is very likely that the engineer will apply the same theoretical principles the student will use to analyze and solve the problem. However, the engineer has to do much more. He has to understand factors such as why the crane is needed in the first place, how and where it will be built, how and by whom it will be used. He will also have to consider the temporal aspects of such factors: how the needs and uses will change over time.

Thus, the designing engineer is dealing with a dynamic situation, whereas the problem solving student is dealing with a static one. However, the engineer will also problem solve when he reduces the situation into smaller elements, and *freezes* it into small problems. The synthesis of the solutions to the constituent problems informs the engineer about the design. However, it does not constitute *the* design as there will always be an arbitrary number of ways of freezing and dissecting any given dynamic situation. Therefore, a design situation will always yield an arbitrary number of *satisfycing*[12] designs. R4 formulates the need for this understanding to be incorporated into the design of the experiment.

**R5:** The setting and scenario of the design experiment should allow for the insertion of control elements associated with the hypotheses, without overconstraining the activity (i.e., quasi-control as opposed to tight control).

I extend the thinking behind R4 in constructing R5, which requires the experiment to employ quasi-control as opposed to tight control when introducing control elements. Clearly, control elements are needed if the experiment is to qualify as an "intervention." However, the point I would like to make is that the nature of the control elements, and hence the amount of control the experimenter has over the experiment, has implications on the nature of the activity that will take place in the experiment.

More specifically, tightly controlled experiments utilize interventions and scenarios which aim to test a specific phenomenon, and that, in doing so, they inevitably promote something other than designing—often problem solving—since they force the scenario to point only at the phenomenon, and the activity to revolve around a specific issue, which is often rather quickly labeled as "the problem." However, as I have argued for earlier, designing never revolves around a singular issue, or, for that matter, a problem.

---

[12] Term borrowed from Simon [Simon 1981].

Therefore, tightly controlled design experiments fail to simulate realistic design situations, and do not promote design activity.

**R6:** The design experiment should facilitate the testing of all hypotheses in a single experiment.

R6 requires the design experiment to facilitate the testing of all hypotheses in a single experiment. There are two reasons for aiming for that. The first one is pragmatic: being able to test all hypotheses in a single session significantly minimizes the effort required to execute the experiment from a logistical point of view as well as the effort to analyze the results from an analytical point of view. The second one is related to the distinction between problem solving and designing I am making: if the hypotheses are tested individually in separate sessions, the activity runs the risk of being reduced to fragmented episodes of problem solving, and R4 and R5 cannot be met.

On the other hand, testing all of the hypotheses in a single session can make it difficult to distinguish the phenomena associated with the hypotheses from each other as they might, and most likely would, be occurring simultaneously. However, I believe that I have minimized that risk by requiring the development of clear definitions and metrics for the phenomena in R1, R2 and R3.

**R7:** The data collection methods used in the design experiment should result in data that can be analyzed qualitatively as well as quantitatively.

R7 ensures that the data generated from the experiments will lend itself to the analysis techniques that are necessary for testing H1, H2, and H3. Judging from the nature of the phenomena under investigation, it is clear that testing H1 relies more on qualitative techniques, whereas H2 and H3 rely more on quantitative techniques.

The two techniques are fundamentally different in the sense that they allow for the tracking and measurement of variables of different natures. In empirical design research, quantitative techniques require precision in identifying localized phenomenon and repeatability of observation of a given data set in order to account for data variables that can be quantified, whereas qualitative techniques require bandwidth of observation in order to capture multiple aspects of activity and account for the relationships between qualitative data variables and other related phenomenon.

It is necessary to point out that this distinction is not necessarily analogous to the distinction I made between ethnographic and audiovisual data collection methods in sections 4.3 and 4.4.1. Even though data generated by audiovisual data collection methods are likely to lend themselves to quantitative analysis techniques naturally, it can still be analyzed with qualitative techniques. Similarly, even though data generated by ethnographic data collection methods is likely to lend itself to qualitative analysis techniques naturally, it can still be analyzed with quantitative techniques. In other words, the choice of analysis method is not directly contingent on the data collection method used.

The choices depend on the specifics of the research project and the nature of the data variables. For instance, when doing field research in order to generate hypothesis, as I have argued for in Sections 4.3 and 4.4.1, it is desirable to use both data collection methods mainly in conjunction with qualitative analysis techniques. When testing hypothesis in the laboratory that require the tracking of qualitative as well as quantitative data variables—as is the case with the experiment discussed in this chapter—it is more desirable (and pragmatic) to use the audiovisual data collection method in conjunction quantitative and qualitative analysis techniques.

## 5.2 Addressing the Requirements

In this section, I will address the requirements and propose ways of meeting them in the design experiment.

### 5.2.1 Defining the Phenomena Outlined in the Hypotheses: The Data Analysis Framework

Developing working definitions for the phenomenon outlined in the hypothesis—question timing (hence frequency), question type, design phase and design team performance—results in an analysis framework for processing the data that will be collected from the design experiment, and addresses R2.

#### 5.2.1.1 Question Definition and Type

In Section 3.1, I defined questions, in the context of this study, as being *verbal utterances related to the design tasks at hand which demand explicit verbal and/or nonverbal responses.* In other words, I take questions to be verbal utterances that demand verbal or nonverbal responses. A response constitutes an answer if it has been solicited by the person whose utterance triggered it—responses which were not explicitly solicited do not constitute answers.

The categories of the taxonomy I proposed in Section 3.5.2 can serve as a categorization scheme to determine question type. The final version of the framework, which I based on Lehnert's original question categories, has 22 conceptual question categories—including Graesser's 4 of 5[13], and my 5 additional categories. Therefore, each identified question can be classified as one of the 22 categories during the analysis.

There is a second method of classification that can be achieved by applying the distinction I also made in Section 3.5.2 between questions that reflect convergent thinking and divergent thinking, which would collapse the 22 categories into 3 conceptual classes: Deep Reasoning Questions, Generative Design Questions, and other (Figure 5-2).

| | Category | Example | |
|---|---|---|---|
| | **Request** | Can you hand me the wheel? | |
| | **Verification** | Did John leave? | |
| | **Disjunctive** | Was John or Mary here? | |
| | **Concept Completion** | What did Mary eat? | |
| | **Feature Specification** | What material is the wheel made of? | |
| | **Quantification** | How many wheels do we have? | |
| | **Definition** | What is a pneumatic robot? | |
| | **Example** | What are some flying insects? | |
| | **Comparison** | Does the small wheel spin faster? | |
| | **Judgemental** | Which design do you want to use? | |
| **Deep Reasoning Question (DRQ)** | **Interpretation** | Will it slip a lot? | **Convergent Thinking** |
| | **Procedural** | How does a clock work? | |
| | **Causal Antecedent** | Why is it spinning faster? | |
| | **Causal Consequence** | What happened when you pressed it? | |
| | **Rationale/Function** | What are the magnets used for? | |
| | **Expectational** | Why is the wheel not spinning? | |
| | **Enablement** | What did they need to attach the wheel? | |
| **Generative Design Question (GDQ)** | Enablement | What allows you to measure distance? | **Divergent Thinking** |
| | Method Generation | How can we keep it from slipping? | |
| | Proposal/Negotiation | Can we use a wheel instead of a pulley? | |
| | Scenario Creation | What if the device was used on a child? | |
| | Ideation | What can we do with magnets? | |

Figure 5-2. A conceptual framework of questions based on Lehnert's taxonomy—including 4 of the 5 categories of Graesser, and 5 additional categories of Eris. Graesser has termed the Deep Reasoning class. Eris has constructed and termed the Generative Design Questions class, and proposed the Convergent-Divergent Thinking distinction.

Clearly, the second method is simpler, and yet, just as meaningful as the first one. Perhaps, it is even more powerful. The detail the first method can provide would play a descriptive function, whereas the higher level understanding the second method can provide would facilitate the testing of the hypothesis.

### 5.2.1.2 Questioning Rate

In order to determine the question asking rate of the design teams in the experiment, each identified question can be time stamped. The beginning of the verbal utterance that constitutes a question can be taken as the temporal pointer. The rate can be calculated by counting the number of questions that are asked in one hour, and reported as

---

[13] I did not consider the "Assertion" category Graesser proposes to be a question since the working definition of a question I use in this study requires a question to demand an explicit response. An assertion does not

questions asked per hour. In order to maintain a level of consistency in the analysis, the videotapes of the experiment should be time stamped while recording. That would ensure the existence of a single canonical temporal reference, and free the analysis from device and user dependant variations. The technical aspects of video recording and replay will be discussed in detail in Section 5.2.4.

#### 5.2.1.3 Design Phase and Process

A design phase can be thought of as a distinct interval in a design project during which functionally similar tasks take place. Conceptually, design phases can be thought as principle constituents of design processes. Design researchers mostly agree on the existence of three such phases even though the vocabulary they use to express them differs: conceptualization, implementation, and assessment. Conceptualization involves tasks geared toward need finding, requirements definition and idea generation, implementation toward specification generation, and assessment toward technical and user testing.

However, design teams do not necessarily execute these phases in that order, or sequentially, nor do they execute them only once. Research in industry shows that, in real-life product development situations, teams go through design phases in varying durations, sequences, and iterations [Hales 1987, McGown 1999], suggesting that teams have unique design processes. The uniqueness of design processes of teams might be dependent on the physical and cultural environment the project takes place in, personalities of the team members, elements related to the nature of the project such as project duration, etc.

H1 postulates that the differences in the design process of a team will be reflected in the types of questions the team asks and the rate it asks them at. In light of the above discussion on design phases and their relationship to design processes, it is possible to test that claim by:

1) Monitoring the design process of a team and observing if specific question asking rates and question types are associated with each design phase.

---

necessarily seek a response.

2) Comparing the overall understanding of a team's design process gained from observing the whole session—or from viewing the audiovisual data collected during a design session—with the understanding gained from considering only the frequency, type and content of the questions asked.

### 5.2.1.4 Design Performance Metrics

Using established performance metrics as a benchmark would enable the testing of the phenomena specified in H2, i.e., the relationship between the frequency of questions and design team performance. In other words, the metric under consideration, question asking, needs to be cross-validated with one or more proven established metric.

Before identifying benchmark metrics for cross-validation with the proposed metric, it is useful to classify design performance metrics into two categories based on the nature of the phenomenon they evaluate: design performance metrics can be based on observable phenomena that occur within design activity, or they can be based on the outcome of design activity—the resulting design or prototype. That distinction deems activity-based metrics as being "internal," and outcome-based metrics as being "external."

Also, it is necessary to note that when measuring performance, I consider the performance of design teams as opposed to the performance of individual designers. There are two reasons behind that decision: as I discussed in detail in Section 2.2, there is agreement within the field that design is a socially mediated activity, and therefore, it should be studied as such when possible. Secondly, when designers work in teams, their questioning behavior is much more explicit since questions are a natural part of the team communication. The implication is that it would be very difficult, and even irrelevant, to attempt to apply an internal or an external metric in judging the performance of an individual team member within the context of the design activity of a team.

The significance and accuracy of the two types of design performance metrics depends on the context they are being used in. Since internal metrics focus on design activity, it is most appropriate to use them to judge the quality of the processes of design teams. And since external metrics focus on products of design activities, it is most appropriate to use them to judge the quality of the resulting designs—the quality of resulting objects such as physical prototypes, or of conceptual representations such as specifications of systems. However, that appropriation does not imply that internal metrics and external metrics are independent, as the outcome of the design activity is, by definition, contingent on itself (Figure 5-3). Therefore, fundamentally, internal metrics can be assumed to be in agreement with external metrics[14].

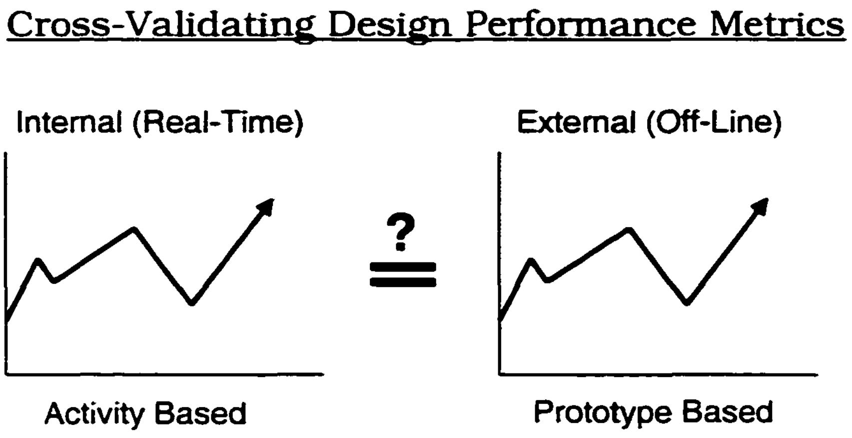


Figure 5-3. The metric under consideration, question asking, needs to be cross-validated with one or more proven metrics. I classify activity based metrics as being "Internal," and outcome based metrics as being "External." Fundamentally, the two are assumed to be in agreement since the outcome of the design activity is, by definition, contingent on itself.

The proposed metric, question asking, is activity based, and, therefore, internal. When cross-validating it, it will be compared with the following two benchmark metrics that are external:

**M1:** The degree of satisfying a set of explicit design requirements by the design.

**M2:** Experts subjectively judging the design (they are asked to assume they are potential consumers about to make a purchasing choice between the available designs).

---

[14] This claim, although somewhat intuitive, will be revisited and tested when analyzing the data as a secondary level hypothesis.

#### 5.2.1.4.1 Benchmark Metric One: Satisfying Given Design Requirements

**M1:** The degree of satisfying a set of explicit design requirements by the design.

M1 is a function of how well the design produced by a team meets its design requirements. In the context of the experiment, this metric is appropriate since a minimal number of basic requirements will be provided to the design teams by the experimenter. (The subjects are still expected to challenge and define most of the requirements, but due to time constraints and for the purposes of providing the necessary experimental structure, they will not get to redefine all of them.) However, in a realistic design situation, all of the basic requirements will most likely be negotiated to some degree by the designers.

#### 5.2.1.4.2 Benchmark Metric Two: Experts Judging the Artifact

**M2:** Experts subjectively judging the design (they are asked to assume they are potential consumers about to make a purchasing choice between the available designs).

M2 implies that design performance is, in the case of a multi-user product, a function of how much demand the design ultimately generates from users, which is essentially a measure of how well design requirements might map onto user demands. The experts will be provided with—apart from the prototypes of the design—basic information about the design such as pricing and another piece of key standard performance information such as device speed, which an average consumer can learn about the product by glancing at the basic specifications listed on the packaging of the product. The expert will then be expected to interact with the prototype, and reach a judgement based on his/her experience with the prototype and standard information he/she has been provided with.

### 5.2.2 Intervening in order to Control Access to Hardware

One way of promoting a clear distinction between designers working with and without hardware in the experiment is to regulate the teams' access to prototyping hardware. More specifically, a certain number of teams will be provided with the hardware at the start of the experiment, and the rest of the teams will be made to work without hardware until midway through the experiment. The teams that start the exercise with hardware will constitute the *control* group, and the teams that receive the hardware midway through will constitute the *test* group since the intervention is the delayed introduction of hardware into the interaction.

The expectation is that the teams receiving the hardware midway through the experiment, in the absence of hardware, will operate at a more conceptual level, and when introduced to the hardware, will switch to operating at a more concrete level. The teams with access to the hardware from the beginning can serve as a natural benchmark for comparison. Thus, the timing of the introduction of the prototyping hardware is the control variable.

### 5.2.3 Promoting "Design Acitivity" as opposed to "Problem Solving"

R1, R4, R5 and R6 are related; meeting one of them implies a degree of meeting the others. The relationship between them is best expressed in R4, which requires the experiment to promote designing as opposed to problem solving. Therefore, it is appropriate to treat R1, R5 and R6 as subsets of R4.

The most effective way to ensure that the experiment will promote designing as opposed to problem solving is to divide the experiment into the following two key constituents and address them separately: the context in which the exercise takes place, and the scenario it utilizes.

A team-based (social) environment—as opposed to designers individually working on specific aspects of what is being designed—resembling a common contemporary design setting in industry, which requires the subjects to fulfill different organizational functions such as engineering, manufacturing, and marketing, can help establish the appropriate context. This viewpoint is valid since the design of new products almost never entails individual designers working in isolation, and is recognized more and more as an interdisciplinary endeavor. Such a context can facilitate the design teams to display sensitivity to multiple perspectives—a characteristic quality of designing—and keep them from operating in a specific domain.

An "ill-defined," and, thus, open-ended design scenario can be utilized in order to guide the teams in the direction of a functional yet original design. Achieving open-endedness in the design scenario entails defining the endpoint of the design scenario as a direction rather than the comprehension and solution of a specific "problem," The expectation is that the utilization of such a scenario would encourage the teams to negotiate and challenge their goals, and would discourage them from committing to a narrow and unrefined set of goals.

### 5.2.3.1 Employing Quasi-control as opposed to Tight Control

The two ways of addressing the key constituents of the experiment I outlined above, requiring the teams to display sensitivity to multiple perspectives as opposed to forcing them to work in a specific domain. Defining the endpoint of the design scenario as a direction rather than the solution of a specific problem also ensure that the experiment will employ quasi-control as opposed to tight control. In other words, they allow for the insertion of control elements associated with the hypotheses into the experiment without overconstraining the activity.

The analysis framework I presented in Section 5.2.1 also serves as a means to employ quasi-control. The variables associated with the phenomenon that make up the framework occur naturally in design activity, and therefore, can be tracked and measured nonintrusively. The only intrusive control element that can be intrusive, and can result in a high degree of control over the design activity, is the delayed introduction of the prototyping hardware to the test teams. Its effects can be assessed and accounted for by qualitatively comparing the resulting activity of the test teams with the more natural activity of control teams.

### 5.2.3.2 Testing of all Hypotheses in a Single Experiment

The hypotheses I outlined in Section 4.4.3 are compatible with each other in the sense that the nature of the design activity that needs to be observed in order to test them is similar. The hierarchical analytical framework for understanding and measuring design performance I constructed in Section 4.4.4 based on the phenomenon outlined in the hypotheses constitutes evidence for that similarity—the hypotheses build on and complement each other. Therefore, for the sake of constructing an initial design exercise, it can be assumed that there are no foreseeable obstacles to testing all hypotheses in a single experiment.

Also, I believe that the analysis framework I presented in Section 5.2.1 is specific enough to allow me to identify the variables of interest, which might be occurring simultaneously if all hypothesis are tested in a single experiment, and to track them accurately.

#### 5.2.3.3 Promoting Realistic Question Asking

What I mentioned in the preceding parts of this section should, for the most part, ensure that the teams practice realistic question asking processes. In other words, if I can ensure that experiment promotes designing as opposed to problem solving by realizing what I have suggested in this section, it would be plausible for me to assume that it also promotes realistic question asking processes.

#### 5.2.3.4 Limitations to Creating Reaslitic Design Situations in the Laboratory

Attempting to create a realistic design situation in the laboratory certainly has many limitations. At best, such an approach can be treated as a "simulation," which implies that the findings can be significantly strengthened by validation in industry. I will discuss some of those limitations in the next chapter when I evaluate the nature of the design activity the pilot experiments promoted. Here, I will outline two of the fundamental limitations: the duration and the context of design activity that can be promoted in the laboratory.

The duration of a real-life design project in industry can range from weeks to years. The key implication of a design project spanning a long time period is the higher extent of learning that would be experienced by designers. More specifically, in the context of this research, it is very likely that the type of learning which takes place over a long time period directly affects the nature and frequency of the questions asked while designing, and that such effects would not be accounted for in the laboratory.

The same thinking is valid for the context the design activity occurs in; it would be foolish to assume that the context a laboratory experiment can provide for a design situation—no matter how well thought out and complex it may be—is identical to the context of realistic design situations in industry. What I attempted in this section was to ensure that the context the experiment I am designing *resembles* the context of realistic design situations in industry as much as possible, so that the findings would be relevant, and, thus, be worthy of validation in industry.

### 5.2.4 The Design Observatory: A Research Instrument and Methodology for Capturing Design Activity in the Laboratory

As I mentioned in Section 5.1.3, when testing hypothesis in the laboratory that require the tracking of qualitative as well as quantitative data variables, the most appropriate data collection method that allows for the utilization of both quantitative and qualitative analysis techniques, and, hence, meets R7, is the audiovisual data collection method.

Audiovisual data provides the precision for identifying localized phenomenon and the repeatability of observation of a given data set quantitative techniques require in order to account for data variables that can be quantified, as well as the bandwidth of observation qualitative techniques require in order to capture multiple aspects of activity and account for relationships between data variables and other related phenomenon.

Tang proposed an experimental facility configuration that facilitates the collection of audiovisual data from design activity [Tang 1991]. His configuration evolved over the process of conducting eight design experiments. He advocated that it is beneficial to:

- Locate the experimenter in a separate room than the room designers are working in.
- Record multiple views of the design activity.
- Keep the cameras stationary.

The specific configuration of the experimental space Tang constructed is illustrated in Figure 5-5. (Tang's laboratory was temporary, and was dismantled after his dissertation work.)

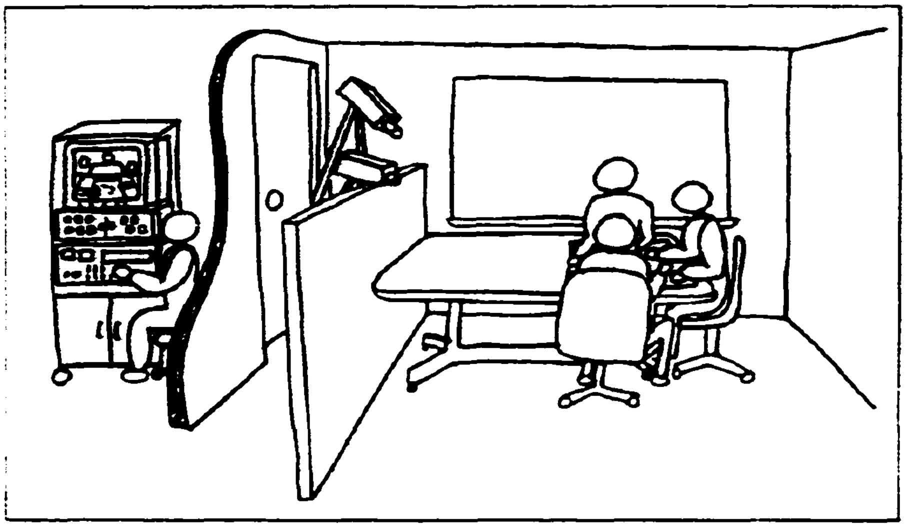


Figure 5-5. Tang's illustration of the practice of the audiovisual data collection method in the laboratory [Tang 1991]. The experimenter is located in a separate room than the room designers are working in. The activity is recorded via multiple stationary cameras.

In order to facilitate the effective collection of audiovisual data, and meet R7, I decided to build a permanent design research laboratory that would be based on and augment Tang's work. Together with my design researcher colleagues Carizossa, Milne, and Mabogunje, I undertook the project in November 2000. The resulting space, named "The Design Observatory," was completed in February 2001.

Similar to Tang's temporary laboratory, the Design Observatory consists of two rooms—one for the designers to work in, and the other for the researcher to monitor the experiment and to collect and process the data. In the design space, there are six cameras, five microphones, a large whiteboard, a work surface, and chairs (Figures 5-6).

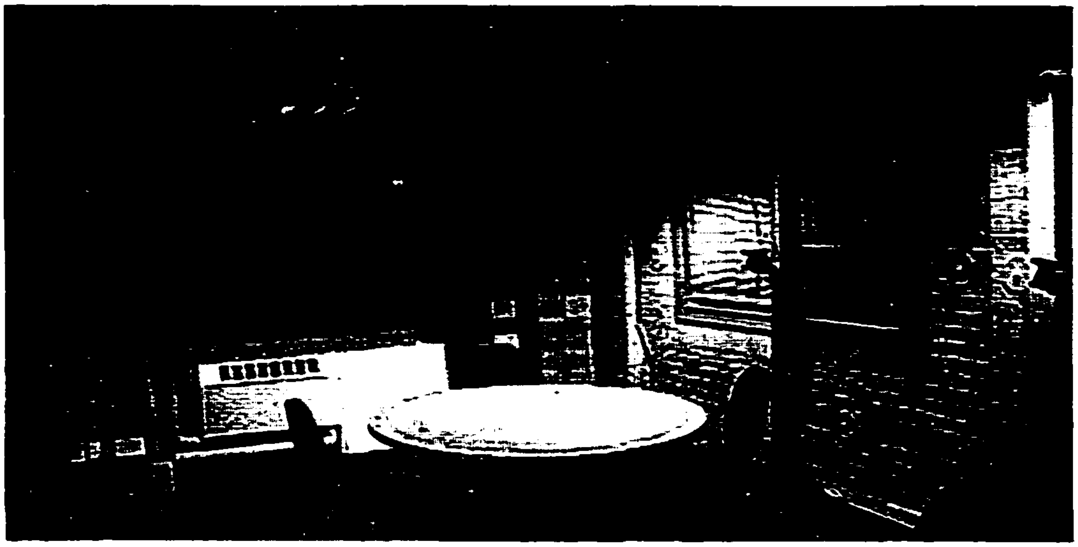


Figure 5-6. The design space of the Design Observatory at the Center for Design Research in Stanford University.

In the data collection and analysis space, there is an equipment rack with personal computers, a video-quad, an audio-mixer, a television and a VCR (Figures 5-7). In order to share the specifications of the Design Observatory with the community and aid other researchers who might be interested in building a similar space, we documented the facility in detail in a publication [Carizossa et. all 2002].

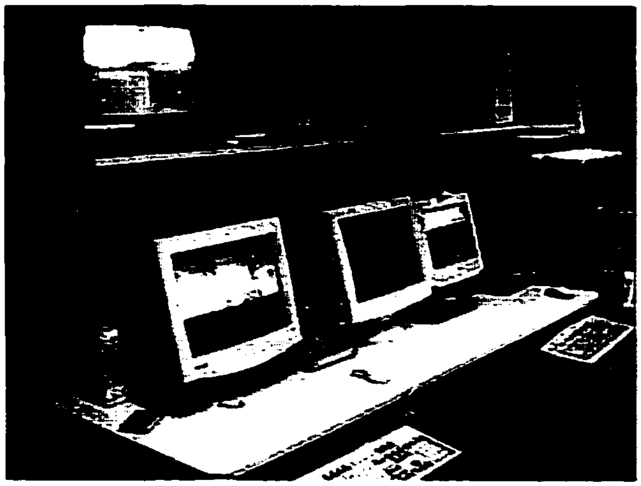


Figure 5-7. The data collection and analysis space of the Design Observatory at the Center for Design Research in Stanford University.

During data collection in a typical design session, the experimenter chooses and orients up to four of the cameras prior to experiment, informs subjects of their confidentiality rights, starts the audiovisual recording instruments, introduces the design exercise, moves to the data collection and analysis space, and monitors the experiment and data recording process from there. The resulting audiovisual data are recorded in split screen format, and if four cameras are used, appear similarly to the sample frame shown in Figure 5-7.

Figure 5-7. A frame from digital video data collected during one of the pilot runs of the design experiment at the Design Observatory.

### 5.2.4.1 On Collecting and Analyzing Digital Audiovisual Data

Technologically, the most significant contribution of the Design Observatory is its digital media capability; the audiovisual data were captured, recorded, and stored in digital format[15]. In that sense, the facility provides a technological enhancement to Tang's paradigm.

In a boarder context, utilizing digital technology to capture design activity is not necessarily a new approach. Researchers developing concurrent and collaborative engineering support tools have been, and still are, experimenting with such technologies to achieve various goals.

---

[15] In order to establish redundancy, audiovisual data is also recorded in analog format with a VCR.

However, utilizing digital technology to analyze data can be seen as a contribution, as it provides new affordances for design researchers. The most significant ones are enhanced audiovisual quality, portability, and potential for indexing of data. High audiovisual quality shortens analysis time and increases precision. Enhanced portability means that data can be shared faster and with a broader audience, which allows for it to be collectively interpreted—inter as well as intra research groups[16]. Enhanced potential for indexing of data can lead to the creation of new cross-referencing methods. As I mentioned in Section 2.3, Yen has already taken advantage of that potential and made an advance in cross-referencing of tacit information with sketching when creating the software tool RECALL [Yen 2000].

## 5.3 Meeting the Requirements: The Pilot Experiment

In the previous section I discussed and specified ways of meeting the seven design requirements for the experiment. The most productive way of integrating those specifications into the initial design for the experiment is to review existing design exercises used by design instructors and researchers that have similar specifications, and adopt one.

The rationale for such an approach is embedded in the nature of designing. Since what I refer to as designing is meant to be complex, it is difficult to predict if a given set of specifications for it will actually produce it. In order to minimize that risk, the most appropriate starting point is to identify an exercise that is known to successfully simulate design activity, and then modify it as necessary. In other words, a convenient way to design a design exercise for my purposes was to *redesign* an existing one with known specifications and consequences similar to the ones that are desired.

With that understanding, I reviewed several existing design exercises. I identified the “Bodiometer Challenge” as a suitable candidate, which was originally created by Mark Cutkosky. In light of the seven requirements, I modified it to the following form, which

---

[16] A research project, known as the Delft protocol analysis, involving collective interpretation of a data set collected from a design experiment was undertaken by Cross and Christiaans [Cross and Christiaans 1996]. However, due to the technological limitations at the time, data could only be shared in analog format.

became the pilot version of the design experiment (for a complete version of the subject instructions, see Appendix A):

The subjects were asked to design and prototype a measurement device called a "bodiometer," which can be moved along male and female body contours to measure their length, with an operating range from 3 to 100 inches. They worked in teams of three, and had 75 minutes to design and construct a prototype from a standard LEGO parts kit that contained a variety of structural and mechanical components, fittings and gears. One group of teams, the control group, was provided with the prototyping materials at the beginning, and the other group of teams, the test group, approximately 35 minutes into the exercise. At the beginning, the test teams received a set of pictures of a representative sample of parts that are in the kit instead of the hardware (for the parts catalog provided to the test teams, see Appendix B). All teams were provided with a set of instructions and a points scheme, outlining how their prototype would be scored once it was constructed. The points scheme accounted for performance topics such as manufacturability, accuracy, cost and aesthetics.

# 6 Conducting and Learning from The Pilot Experiments

The third step of the empirical dimension of this research has two parts. The first part entails evaluating and redesigning the initial version of the experiment. In order to test the initial version, I conducted pilot runs with two design teams, one being the control group, and the other the test group. The pilot runs played a critical role in improving my experimental methodology, deepening my understanding of the nature of questions, and augmenting my hypotheses. These advancements were then reflected in the redesign of the experiment.

In this chapter, I evaluate the implementation of the requirements discussed in the previous chapter in the context of several observations I made while conducting the pilot experiments. In each of the initial three sections, I discuss one of the advancements I mentioned above in detail. In the last section, I summarize the augmented hypotheses.

## 6.1 Improving the Experimental Methodology

In order to improve the experimental methodology, I observed and assessed the conditions during the pilot runs in the context of the four design requirements under the design research experimentation criteria, R4 through R7.

The pilot runs did not reveal any fundamental difficulties in meeting R4 and R5; as intended, the exercise promoted designing rather than problem solving as a whole. The two design teams spent a significant amount of their time and energy in negotiating and redefining the requirements, and explored a variety of different designs. For the most part, their approach did not suggest that they viewed the requirements as "givens," and the outcome of their effort as "the solution." They seemed to be aware that the requirements they were acting on, and the designs they were creating were possibilities. Also, both teams displayed sensitivity to multiple perspectives: they considered user needs, manufacturability and cost issues, and aesthetic values, as well as addressing conceptual and technical issues that were in their domain of mechanical engineering expertise.

The intervention, delaying the introduction of the hardware to the test group, did not seem to break-up the team's workflow and fragment the activity. The team members continued to work without interruption, and did not feel the need to rethink their process when they received the hardware. However, as intended, the intervention affected the activity by promoting them to conceptualize more in the absence of hardware. This observation indicates that the nature of the intervention was balanced and not overly contradictory of the natural design processes of the team.

However, the pilot runs did help me to identify a number of issues related to R4 and R5 that needed to be addressed. The most significant one was the timing of the introduction of the hardware to the test group. At the beginning of the exercise, the test group was informed that they would be receiving the hardware 35 minutes after the start of the exercise. For approximately the first 10 minutes, they seemed cognizant of that milestone, but once they got into the exercise and focused solely on designing, they lost track of it. After about 25 minutes, they stopped conceptualizing and indicated that they were ready for the hardware—the point being that if they had not lost track of the milestone, they might have paced themselves accordingly. I saw no reason to force them to conceptualize for another 10 minutes since insisting on the intervention to occur that way might have broken up the teams workflow, and decided the give them the hardware earlier than anticipated.

In other words, releasing control to the team on that matter helped the team transition smoothly and improved their workflow. Therefore, I decided that giving the test teams the choice of asking for the hardware when they felt ready to proceed to working with hardware, instead of forcing them to conceptualize for a fixed amount of time, would be a better way implementing the intervention.

The pilot runs also revealed that it was necessary to change the structure of the points scheme used for evaluation of the prototype according to M1 in order to avoid a situation where teams which might be inclined to approach the exercise with a problem solving framework might focus solely on optimizing their score (the points scheme is presented in detail in Section 7.1.4.1). The intent of the points scheme was to provide the teams with a sense of what might be important to the potential users of the bodiometer device. However, during the pilot runs, it became clear that when the points scheme was too explicit, it lost its intended function, and, instead, promoted such teams to become immersed in the optimization of the algorithms used for the calculation of their score without considering what they were meant to convey.

In the pilot experiment, points could be earned for satisfying each of the following functional and user requirements with the prototype: accuracy, aesthetics, operation time, number of parts, manufacturing time and design concept. (For a detailed description of the requirements, please see the subject instructions in Appendix A.) The linear algorithms used in the calculations were made explicit in the instructions. For instance, each part used and second elapsed in manufacturing cost the team a fixed number of points. That method of points allocation resulted in an absolute points scale, and both pilot groups spent significant amounts of time attempting to optimize the relationships between the algorithms in order to maximize their score without considering what the scale was meant to suggest.

Therefore, I decided to use a relative points scheme instead, in which points would be assigned based on the rank a prototype achieved among all prototypes in meeting a specific requirement. The teams would not be informed of the performance of other prototypes, and, in the presence of that ambiguity, would be forced to consider the meaning and importance of a specified requirement first as opposed to immersing themselves in calculating the optimal degree of meeting it.

The duration of the exercise proved to be too short for the teams to create a direction for their designs and execute it, as both teams were still negotiating the requirements with

30 minutes remaining. Therefore, I decided to raise the duration of the final version of the exercise to 90 minutes.

Even then, the time limitation had implications. Perhaps, it was the most significant limitation for the experiment, since it is difficult to guarantee that the 90-minute design exercise is indeed a condensed version of a long-term design project. For example, it is possible that the nature of questions asked by designers change after six months of reflection on a design—the taxonomy I use might not even have a category to accommodate such questions. Although I took many steps to ensure that the key characteristics of the questioning behavior of professional designers working on real-life design projects will be replicated in the experiment, I cannot know how successful I have been in achieving that unless I go back to industry and attempt to validate my laboratory findings there. That is the inverse of what I attempt to accomplish in this dissertation, and would constitute a very worthy follow up study on its own.

The pilot runs did not uncover any difficulties in meeting R6, even though testing all hypotheses in the same exercise resulted in the phenomena associated with the hypotheses to occur simultaneously. The definitions I developed for the phenomenon, and for the forms I expected them to manifest themselves in the data, aided me in discriminating the data variables and tracking them independently.

Meeting R7 by utilizing the digital observation and analysis technology I developed proved to be feasible as well. However there were two technical issues that needed to be addressed: limitations in mobile digital storage space and playback bandwidth.

I determined that the computer dedicated to capture and playback the audiovisual data needed to support a minimum data transfer rate of 1000 Kb/s in order to attain reasonable image quality at a resolution of 640 by 480 pixels and mono sound at 11.2 kHz scan frequency. The size of captured video file from a single experiment would be roughly 4 GB. At the time, that posed an issue, as available portable storage devices such as CD-Roms and floppy discs could not store that much data, and most external hard-drives could not support the 1000 Kb/s transfer rate[17]. However, right after the pilot experiments, external hard-drives utilizing the FireWire data transfer protocol, which met the storage size and transfer requirements, became available. That technology made it possible for me to record 15 experiments in a single 60 GB external drive.

Storage technology has been advancing. Today, it is possible to use DVD-R drives in writing digital data to DVDs that can hold upto 4 GB data each. Thus, audiovisual data from a single experiment can be stored on a single disc. That makes the sharing of digitized experiment data rather effortless, as DVDs can be easily replicated and handed off or mailed to a colleague. Also, there are more efficient audiovisual compression protocols available now, which should reduce the 4 GB per experiment storage requirement.

## 6.2 Augmenting the Hypotheses: Discovery Making as Another Internal Performance Metric

In order to refine my hypotheses, I reconsidered them briefly in light of the observations I made during the pilot exercises. Even though the limited sample size of the data generated by the pilot runs did not permit me to draw conclusions, what I observed enabled me to deliberate on their relevance and validity.

When I reconsidered H1, I discovered enough evidence to convince me that it deserves detailed investigation. Paying attention to the nature and timing of questions asked by the two design teams allowed me to gain a comparative understanding of their question asking process. Also, when that understanding was viewed from a broader scope, it seemed to suggest a topographic representation of the design activity that took place.

I also found qualitative as well quantitative preliminary evidence in the data suggesting that, as I postulated in H3, the intervention employed in the experiment affected the questioning behavior of the teams. The test team asked more questions in the absence of prototyping hardware (a 21% increase in the second phase of the experiment),

---

[17] The ability to use portable data storage devices is important since, in Section 5.2.4.1, I argued that one of the main affordances of digital technology is the portability it provides for data.

whereas the control team asked about the same number of questions in each phase (a 5% increase in the second phase of the experiment).

Reconsidering H2 raised two issues regarding M1 and M2, the external benchmark performance metrics I proposed in Section 5.2.1.4. Firstly, as I addressed in Section 6.1.1, it was evident that the points scheme I used to score the prototypes, the method for obtaining M1, required modification. I also realized that even if the points scheme had been sound, comparing the two data points obtained from the pilot runs (M1 results in one performance measurement for team) would not be meaningful.

Secondly, it was also evident that obtaining M2, evaluation of the prototypes by experts, was not feasible at that stage for the same reason; experts comparing and ranking only two prototypes was not particularly insightful as a performance measure. Therefore, in the context of the data generated from the pilot runs, I will not speculate on the relationship between question asking and the benchmark performance metrics.

Recognizing those issues helped me to identify a characteristic limitation associated with external metrics: measuring performance in terms of the outcome of the design activity, the design, means that the measurement is made on a single object, the prototype, regardless of how many different metrics might be employed. For instance, M1 and M2 are different metrics, but they operate on and judge the same prototype[18].

However, internal metrics are not necessarily subjected to the same limitation since the phenomenon associated with an internal metric most likely occurs numerous times within the activity[19], and it is very possible that each occurrence directly or indirectly causes another performance phenomenon. The identification of such a related performance phenomenon might possibly result in another performance metric, which

---

[18] The premise of that argument is that there exists only one "design," and hence, prototype. However, even if the outcome of the design activity is considered to be multiple designs, their numbers can be assumed to be small. It is unrealistic to think 10 prototypes will be produced in a design project. Even though 10 "design concepts" might be created and considered, it is plausible to assume that not more than 3-4 will be implemented in the form of functional prototypes.

[19] If the phenomenon associated with an internal metric does not occur multiple times within the activity, it would be difficult to measure, and attempting to measure it would not yield statistical significance. In other words, it would be meaningless to attempt to establish it as a metric.

would essentially be a surrogate of a surrogate of the principle of interest[20]. Therefore, identifying an additional internal performance phenomenon related to question asking that occurs *within* the activity would give me multiple measurements, and, hence, multiple data points per team, even within the limited data set generated from the pilot experiment.

In order to identify such a performance phenomenon, I rescanned the data from the pilot runs and compared my observations with the observations I made of the paper bicycle team. I found an observation I made regarding the discovery making process of the paper bicycle design team, O2, particularly relevant to what I saw in the pilot run data. O2 states that the paper bicycle design team seemed to discover more when they asked "good" questions. What I observed in the pilot run data was an extension of that observation: the pilot teams seemed to conceptualize more articulate and a higher number of designs when they *discovered* more. Therefore, I decided to consider "discovery making" as another internal performance metric. That constitutes an additional hypothesis, H4, to supplement the three I have listed earlier.

When identifying a discovery within the activity, I looked for instances where the team experienced a realization that lead to a unique and previously unthought of concept, or obstacle, related to the design they were working on. Within the scope of the specific design exercise I used in the experiment, I identified four areas in which such conceptual leaps could occur: measurement concept, readout concept, mechanism concept and obstacle recognition. It is appropriate to note that this method is somewhat similar to judging the effectiveness of a brainstorming session by the number of ideas created. However, discovery making is different from ideation in the sense that it involves a higher and more visible degree of conceptual continuity and progression, and, therefore, is most likely strongly tied to learning.

To summarize, my deliberations on the limitation of utilizing external metrics, and the relevance of identifying another internal performance phenomenon resulted in the addition of the following hypothesis, H4:

---

[20] The "surragote" and "principle" terminology borrowed from Ijiri [Ijiri 1967].

**H4:** There is a strong correlation between the frequency of discoveries made by design teams and design team performance. Hence, discovery making can be taken as a performance metric.

## 6.3 Refining the Hypotheses: Definition of a "Good" Question

In order to deepen my understanding of the nature of questions, I reassessed the principles and the structure of the taxonomy by testing it as a coding scheme for the questions raised in the pilot runs. I also advanced the discussion I initiated in the previous section on discovery making, and developed a better understanding of what a "good" question might be.

Surprisingly, and perhaps because the principle of the taxonomy is sound, when I attempted to code the questions I identified in the pilot exercises with the taxonomy I presented in Section 3.5, I did not experience prolonged indecision in assigning categories to any of the questions—provided I had enough time for each assignment and did not lose focus by coding more than 20 questions in a row without resting. As an alternative coding method, I classified the questions with the three encompassing question classes I discussed in Section 5.2.1.1: Graesser's DRQs categories, the GDQ categories I constructed, and "other." Utilizing the more encompassing question classes in coding resulted in a much faster process, and did not result in any significant ambiguities in categorization either.

When I used the taxonomy to code the data, all 22 categories received multiple hits[21]. The distribution was not even (and it does not need to be). Lower order questions were more frequent. The important observation is that, in employing either coding scheme, I felt the need to utilize all of the categories, and did not encounter any questions that I could not classify[22].

---

[21] During the analysis of the data collected from the pilot experiments, I acted as the only coder.

The rationale I developed in constructing H4 motivates and aids me in defining what a "good" question might be; I reiterated my observations regarding the paper bicycle design team appearing to discover more when they asked "good" questions, and the pilot teams appearing to conceptualize more articulate and a higher number of designs when they discovered more. Thus, it is natural to ask the following question: What was distinct about the questions the paper bicycle and pilot design teams asked that might have been related to them discovering more?

In order to answer that question, I focused on the instances of discovery making in the data and identified the questions occurring before them—the assumption being that "good" questions are associated with discovery making. A significant part of the questions I identified were a combination of DRQs and GDQs.

That observation is in agreement with Graesser's rationale for assigning a higher degree of importance to DRQs than the other types of questions. As discussed in Section 3.4, Graesser argues that DRQs are associated with achieving the higher level goals listed in Bloom's taxonomy of educational objectives [Bloom 1956], and empirically demonstrates that the asking of DRQs are correlated with learning performance in tutoring situations. However, the tutoring situations Graesser studied tend not to promote the type of learning that occurs in a design context. Therefore, I wondered if GDQs might also be correlated to performance, but within a design context.

That is not to say that I assumed that the importance Graesser assigned to DRQs was invalid in a design context. On the contrary, I saw no reason to believe their occurrence would not contribute to a correlation with performance in a design context as well. I postulated that, in order to account for a correlation between question asking and design performance, GDQs needed to be considered as a necessary addition to DRQs, and that they needed to be treated as a pair.

---

[22] In the case of the compelete taxonomy consisting of 22 categories, that was partially because some of the categories such as the "Proposal/Negotiation" category are rather broad conceptually.

That consideration can be best studied if it is translated into a hypothesis. The most appropriate way to do so is to incorporate its premise into the existing hypothesis regarding the relationship between question asking and performance, H2, by focusing on the DRQ-GDQ pair as opposed to all types of questions, and testing a correlation between them and design team performance.

Therefore, I decided to modify H2 to the following form:

**H2:** Two classes of questions, termed Deep Reasoning and Generative Design questions, are related to design team performance. Their frequency of occurrence correlates strongly with design team performance, and can be taken as a performance metric.

This modified hypothesis, together with the new hypothesis presented in the previous section, reflect two of the three elements of what a "good" question might be in a design context. To summarize, the three elements of a good question are its:

1) Semantic structure
2) Consequences
3) Content

Throughout this dissertation, I argue that two classes of questions, DRQs and GDQs, reflect the semantic structure of good questions, and that the posing of good questions often lead to conceptual leaps, or rather, to discoveries. However, the formulation of the third element, the one I have not addressed, is problematic because it is strongly associated with the context the question is posed in.

## 6.4 The Augmented Hypotheses

To summarize, the final states of the hypotheses are the following:

**H1:** Question timing and type are descriptive characteristics of design cognition and process. When the set of questions a design team asks during a design project is considered as a whole, the timing and nature of those questions point at the fundamentals of the knowledge and rationale the team uses for breaking down and structuring the project into design phases. Question timing and type are informative enough to serve as a roadmap to the design thinking and process of the team.

**H2:** Two classes of questions, termed Deep Reasoning and Generative Design questions, are related to design team performance. Their frequency of occurrence correlates strongly with design team performance, and can be taken as a performance metric.

**H3:** Question asking behavior of design teams is influenced by their access to hardware. The types of questions design teams ask change when they transition from working in the absence of hardware to working with hardware.

**H4:** There is a strong correlation between the frequency of discoveries made by design teams and design team performance. Hence, discovery making can be taken as a performance metric.

# 7 Conducting The Redesigned Experiment: Putting the Question Asking aspect of Design Cognition under the Microscope

The second part of the third empirical step of this research involves conducting the final version of the experiment and analyzing the data. After redesigning the exercise to be used in the experiment, and improving the experimental methodology by reflecting on the pilot experiments, I conducted the final version of the exercise with twelve design teams, consisting of three designers each. I then analyzed the resulting data according to the frameworks presented in Chapters 3 and 5 in order to test the four hypotheses.

In the first section of this chapter, I discuss the data collection and analysis procedures. In the second section, I present my analysis of the data. In the third section, I revisit the hypotheses in light of the results of my analysis.

## 7.1 Data Collection and Analysis Procedures

In this section, I derive data analysis procedures from the analysis frameworks I presented in sections 5.2.1 and 6.3, and complement the discussion on experimental methodology I presented in sections 5.2.2 and 5.2.3 with the specification of experimental procedures.

### 7.1.1 Subject Recruitment and Design Team Composition

Subjects were recruited in person and via group email messages. The only prerequisites for being a subject in the experiment was to be a currently enrolled student in a mechanical engineering graduate program at Stanford University, and not to have any prior knowledge of the "Bodiometer" design exercise. The first twelve subjects making up the first four teams were volunteers. The rest were paid $20.00 each for their participation.

Subjects were encouraged to apply in groups of three. The ones who did so were treated as a design team. Four teams were formed that way. Two of those teams were assigned to the test group, and two to the control group. Apart from those four, assignment of teams to experimental groups was performed randomly. There were no guidelines for forming the other eight teams, and, for those eight teams, assignment of subjects to teams was performed randomly. However, the subjects making up seven of the eight teams knew each other—they had either worked as a part of a group in a class or on a research project before, or they were a member of the same academic research group. The subjects making up one of the eight teams had not met before.

The subjects were encouraged to form teams with people they knew so that they would be comfortable with expressing themselves and ask questions freely. It is true that forming teams in such a way did not necessarily promote homogeneity between them, but from the viewpoint of measuring team performance, that was not required.

### 7.1.2 Experimental Procedure

Just prior to the beginning of the experiment, the design team members were introduced to the functionality of the Design Observatory in order to make them comfortable with the setting. Each audiovisual recording device in the design space was explicitly identified, and the procedures for handling captured data were explained. Consent forms were handed out, and the team members were allowed the necessary time to read and understand the material. Upon receiving consent from all three members, audiovisual recording was started, and subject instructions explaining the design exercise were handed out according to the experimental group the team was placed in—test or control. (For subject instructions, see Appendix A.) The experimenter stayed with the team and answered any questions until all team members indicated that they understood the instructions.

The experimenter then moved next door to the data collection space of the Design Observatory, monitored the activity from there by observing the feed coming into the digital recording equipment from the cameras and microphone in the design space. Before leaving the design space, the team members were told to feel free to say "Question" and wait for the experimenter if they had any questions about the design exercise. In the case the team opted to ask a question, the experimenter quickly stepped into the design space, answered the question, and returned to the data collection space.

Teams in both experimental groups were notified 30 and 10 minutes before the end of the full 90 minutes. Teams in the test group were given the freedom to decide when to stop conceptualizing and start interacting with the prototyping hardware. However, the prototyping hardware was introduced to the test teams even if they had not asked for it if 35 minutes had elapsed. At the beginning of the experiment, the control teams were provided with the prototyping hardware, whereas the test teams were provided with a document containing 15 photographs documenting all part types instead. (For the parts catalog provided to the test teams, see Appendix B.)

The hardware consisted of the "Lego Technic Star Wars Episode I Battle Droid" kit (Lego kit number 8001), which had 328 prefabricated structural and mechanical components, fittings, and gears. Each team was provided with a new unopened box containing the kit as well as the original manual with instructions for constructing the Star Wars Battle Droid.

At the end of the 90 minutes, the team members were given a final warning, and asked to conclude their work. Once the exercise was over, they were asked to identify their prototype and to explain how it worked. They were then provided with another Lego Kit, and asked to identify and prepare the parts their prototype was made of. When they were ready, they were asked to construct a device identical to their original prototype from the parts from the second kit. There was no limit on the number of team members who could participate in the construction of the second device, and they were allowed to use the original prototype as a guiding model. The construction process was timed and recorded as the "Manufacturing Time." All audiovisual recording equipment were then turned off.

### 7.1.3 Transcription

The first two of the twelve experiments were completely transcribed. The speaker, time stamps marking the start and end of the utterance, the utterance itself, and any comments outlining relevant behaviors or circumstances not directly reflected in the utterance, were documented on the transcript. (For a sample segment from the transcript of Team 1, see Appendix C.) Inaudible utterances were clearly marked as such. For reasons I will discuss Section 7.1.5, the rest of experiments were not transcribed.

### 7.1.4 Scoring and Judging the Prototypes

As discussed in section 5.2.1.4, the prototypes constructed by the teams were evaluated according to two external benchmark performance metrics.

#### 7.1.4.1 Scoring the Prototypes according to M1

The first benchmark metric, M1, was a function of how well the prototypes met the provided core design requirements, which were aesthetics, measurement speed and accuracy, manufacturing time, number of parts and measurement display interface.

A combination of the potential cost and sales of the prototype determined the overall team score. The final score for each team was computed by using the following equations:

Score = Sales – Cost, where;
Sales = Design Concept + Aesthetics + Measure Time – Error
Cost = Number of Parts + Manufacturing Time

Teams received a score under each category according to the following rules:

**Error** was scored (10 points for 1 inch of error) as the sum of the absolute values of the differences between the two team measurements and the official measurement where:
Team-measurement = Handweb + Head Circumference
Error = Absolute Value [(team measurement)-(official measurement)]

**Measure Time** was the combined time it took for the experimenter to make the two measurements. Points were handed out in the following way: fastest = 20, next fastest = 15, third fastest = 10, 4th fastest = 5, 5th fastest = 3, 6th fastest = 2, 7th fastest=1, and slowest =0.

**Number of Parts** was the total number of parts used in the prototype. Points were handed out in the following way: highest = 20, 2nd highest = 15, 3rd highest = 10, 4th highest = 5, 5th highest = 3, 6th highest = 2, 7th highest =1, and lowest =0.

**Manufacturing Time** was the time it took the team to rebuild the prototype from an identical and new parts kit after the main part of the experiment was over. Points were handed out in the following way: highest = 20, 2nd highest = 15, 3rd highest = 10, 4th highest = 5, 5th highest = 3, 6th highest = 2, 7th highest =1, and lowest =0.

**Design Concept** was a 30-50 point bonus for a design that provided an instrumented readout. Instrumented readout was any method which allowed the user to "read off" a measurement by simply looking at the device without making any calculations or looking at any value tables.

**Aesthetics** was a subjective category (0-10 points), computed by averaging the scores handed out by 3 judges. Opinions were based solely on the prototype. Visual and intellectual aesthetics were the main considerations[23].

### 7.1.4.2 Judging the Prototypes According to M2

The second external benchmark metric, M2, entailed three experts subjectively judging the prototype. All three experts were professors in the Design Division of the Mechanical Engineering Department at Stanford University.

The experts were provided with—apart from the prototypes—two pieces of basic information about the design: the cost and measurement speed. It was assumed that those key pieces of information about the product would be made available to an

---

[23] Design concept and aesthetics points were assigned subjectively by two professors and the experimenter. The averages were of the three scores were used in the final calculation.

average consumer in the form of basic specifications printed on the packaging of the product. The experts then briefly (5-10 minutes) interacted with the prototypes, and reached a judgement by ranking them against each other.

### 7.1.5 Question Identification and Logging

Initially, questions were identified from the transcripts by utilizing the working definition of a question presented in Section 3.1. However, there were two significant difficulties with analyzing the questions from the transcripts: the lack of context for identifying and categorizing the questions. When attempting to identify questions from the transcripts, the grammar used for posing questions in speech was often misleading. Many of the utterances that conceptually constituted questions were not grammatically structured as questions, and therefore, could not be properly identified. For instance, it was difficult to determine if the utterance "This gear attached to the long rod" constituted a question or not by simply analyzing the transcript because transcripts did not provide the necessary context. However, audiovisual data did.

Even if a question was correctly identified from a transcript, when attempting to categorize it, it was difficult to determine the category it should be assigned to due the lack of contextual information. For instance, it was almost impossible to determine if the question "Can you move the wheel?" should be assigned to the Request or to the Proposal category by analyzing the transcript. Furthermore, in some cases, it was difficult to make such judgements even from the audiovisual data, and a 2-3 minute interval during which the question had been posed had to be viewed repeatedly three or four times for clarification.

After attempting to analyze the first two experiments from transcripts, it was evident that transcripts yielded much less insight compared to audiovisual data. Also, transcripts were not cost-effective, as it took approximately 15 hours to transcribe 1 hour of audiovisual data. Therefore, the other ten experiments were not transcribed, and all experiments were studied primarily by analyzing the audiovisual data.

All identified questions were logged on a spreadsheet together with the time stamps marking the start of each question, and the coded identity of the team member asking the question (A, B, or C). The spreadsheet assigned each question a sequential number. Each team member was also assigned a sequential number for each question he/she asked. The spreadsheet calculated the time interval between questions (delta t in seconds), and the overall question asking rate of the team until that point (QAR in questions/hour). Once a category of the question was determined, the corresponding number for it was also recorded (Cat). The initial section of a sample spreadsheet can be seen in Figure 7-1.

| Time | Question | A | B | C | Delta t | QAR | Cat | Question |
|---|---|---|---|---|---|---|---|---|
| 0 | | | | | | | | [START SESSION] |
| 556 | 1 | 1 | | | 556 | 6.5 | 1 | Hand web is from here to around there? [E] |
| 591 | 2 | 2 | | | 35 | 12.2 | 1 | These measurements, could it be like in legos [units]? [E] |
| 600 | 3 | 3 | | | 9 | 18.0 | 1 | Do we know the length of these legos? [E] |
| 626 | 4 | | 1 | | 26 | 23.0 | 22 | Do you guys mind if I take these [parts list] apart? |
| 0 | | | | | | | | [START EXERCISE] |
| 4 | 1 | | 1 | | 4 | 900.0 | 1 | So, we're doing phase one? |
| 89 | 2 | | 2 | | 85 | 80.9 | 22 | Why don't we make sure we know how readout's going to be graded? |
| 111 | 3 | | 3 | | 22 | 97.3 | 1 | We basically need to measure the perimeter of the contour, right? |
| 160 | 4 | 1 | | | 49 | 90.0 | 18 | Does it have to have multiple linkages? |
| 176 | 5 | | 4 | | 16 | 102.3 | 22 | We'll write it down as a possible idea, right? |
| 179 | 6 | | 5 | | 3 | 120.7 | 6 | What do you call that? |
| 181 | 7 | 2 | | | 2 | 139.2 | 1 | That would be a really simple idea--one piece, right? |
| 201 | 8 | | 6 | | 20 | 143.3 | 1 | And measure how many revolutions? |
| 208 | 9 | 3 | | | 7 | 155.8 | 18 | Or, you could just have a string of legos connected like a linkage? |
| 225 | 10 | | | 1 | 17 | 160.0 | 1 | Do you know what I'm saying? |
| 262 | 11 | | 7 | | 37 | 151.1 | 6 | What do you mean flipping over? |
| 324 | 12 | | 8 | | 62 | 133.3 | 1 | Were you thinking about a one that you'd put together? |
| 348 | 13 | | 9 | | 24 | 134.5 | 3 | What do you call that thing? |
| 354 | 14 | | 10 | | 6 | 142.4 | 1 | And you keep count? |
| 357 | 15 | 4 | | | 3 | 151.3 | 22 | Can I draw something like that just to see if you could X? |
| 384 | 16 | | 11 | | 27 | 150.0 | 1 | That was the first idea, right? |
| 400 | 17 | | 12 | | 16 | 153.0 | 21 | Any more brainstorming ideas? |
| 415 | 18 | | 13 | | 15 | 156.1 | 1 | Is it a requirement that it automatically has to give you a value? |
| 439 | 19 | | 14 | | 24 | 155.8 | 9 | I wonder if this would count though, just wrap it around and read it off? |
| 510 | 20 | 5 | | | 71 | 141.2 | 10 | Do you think that might be more precise? |
| 522 | 21 | 6 | | | 12 | 144.8 | 6 | What's error? |
| 527 | 22 | | 15 | | 5 | 150.3 | 1 | It seems like, it also needs to be long enough to go around your head, right? |
| 557 | 23 | | 16 | | 30 | 148.7 | 1 | Is that what you're saying? |
| 569 | 24 | | 17 | | 12 | 151.8 | 1 | This is 11 inches, right? |
| 609 | 25 | | 18 | | 40 | 147.8 | 6 | What's X diameter? |
| 614 | 26 | | | 2 | 5 | 152.4 | 1 | Your fingers are about 3 inches long, right? |

Figure 7-1. The initial section of the spreadsheet where questions asked by Team 12 during the design exercise were logged.

## 7.1.6 Question Categorization

All identified questions were coded according to the categories of the taxonomy of questions presented in Section 5.2.1.1. There were two critical issues associated with the coding process: as I mentioned in the previous section, in certain cases, it was difficult to comprehend the context of a question even after viewing the portion of the audiovisual data the question occurred in several times; also, when the context was determined, the dependencies between some of the question categories added a second degree of ambiguity that needed to be resolved.

In order to decipher the context in which a question was posed, at a minimum, it was necessary to pay specific attention to, and to interpret the motivation of, the questioner for asking the question, the general direction of the design activity, the present state of the prototype or sketch or any other representation that was being referenced, and any prior exchanges that might have taken place within the group building up to the asking of the question.

Ambiguity resulting from the inherent dependencies between some of the question categories in the taxonomy was resolved by identifying all question category principles applicable to the question under consideration, and prioritizing them in order of intent. In general, it can be assumed that the higher order question categories (in Figure 5-2, categories listed at the bottom are of higher order than the categories above them) are conceptually closer to what the questioner intended, and, therefore, of higher rank. Therefore, when a question is conceptually in agreement with the defining principles of multiple categories, it should be assigned to the category with the highest rank. For instance, most lower order questions are "Verification" questions, and most DRQs and GDQs are "Judgmental" questions to some degree. According to the guideline I presented here, questions were coded as verification questions only if they could not be coded as belonging to another category. Similarly, DRQ and GDQ categories had priority over the Judgmental or the other categories[24].

Reliability testing was done in order to cross-validate the question identification and categorization processes. Two doctoral candidates, a design researcher and a social scientist, served as coders in the cross-validation process. They were not related to this research, and had experience with video interaction analysis and coding. Abiding by the working definition of a question presented in Section 3.1, the social scientist was

---

[24] Graesser also recognized that the version of the taxonomy he used in categorizing questions could be used as a monothetic or polythetic scheme. He observed overlaps between the verification category and other categories, and between DRQ categories and other categories. He argued for a similar rank hierarcy to the one I did based on a slightly different rationale, and opted to use a monothetic scheme for data analysis as well [Graesser 1994].

exposed to 50 questions which had been raised by two different teams in two continuos intervals. 14 of the 50 questions were either DRQs or GDQs. Cross-validating her designations with mine resulted in a reliability score of 0.98. When coding the questions according to the 22 categories, the reliability score was 0.90 (4 of the 5 disagreements were related to questions which I had assigned to specific DRQ or GDQ categories). The reliability score was 0.98 when she coded according to the 3 question classes only.

Since the social scientist did not experience any difficulty in categorizing questions which I had assigned to categories other than the DRQ and GDQ categories, the design researcher was asked to code questions which I had identified as being DRQs or GDQs only. He was exposed to 50 DRQ and GDQ questions which had been raised by three different teams in five distinct continuos intervals, and achieved a reliability score of 0.92.

### 7.1.7 Discovery Identification and Logging

After all questions were identified and categorized, the audiovisual data were scanned a second time in order to identify the discoveries the design teams made. As mentioned in 6.1.2, within design activity, a discovery was considered to be a realization that led to a unique and previously unthought of concept or obstacle. Each identified discovery was assigned to one of the four discovery categorizes specific to the design exercise utilized in the experiment: measurement concept, readout concept, mechanism concept and obstacle recognition.

For each design team, the categorized discoveries were logged in a spreadsheet together with the time stamp marking instance when the discovery was communicated verbally for the first time within the team, and the coded identity of the team member communicating the discovery (A, B, or C). Since discovery making is a continuos and additive phenomenon, it was not appropriate to assign a specific discovery to a specific team member. The only aspect of discovery making that could be observed with confidence was its initial verbalization and communication.

Each discovery was also labeled with a few descriptive words. The descriptive labels were initially unique to the teams. However, after the discoveries made by all of the teams were logged, conceptual similarities emerged between some of them, and those were merged under a single label. The spreadsheet calculated the time interval between the discovery communications, and the overall discovery making rate of the team. A spreadsheet summarizing the discoveries design team 3 made during the exercise can be seen in Figure 7-2.

| Time | Concept | Readout | Mechanism | Obstacle | A | B | C |
|---|---|---|---|---|---|---|---|
| 114 | | | | Can't fit the measurement piece in between fingers | X | | |
| 126 | Rolling a wheel | | | | | | X |
| 168 | | Wheel flicks the read-out | | | | | X |
| 179 | | | Gears | | | | X |
| 179 | | | Gear reduction | | | | X |
| 179 | | Dial | | | | | X |
| 324 | Series of linkages | | | | | X | |
| 420 | Exrapolate from a standart body part | | | | X | | |
| 495 | Set Lengths, a Rod | | | | | X | |
| 532 | | | | Negotiating sharp angled countours and corners | X | | |
| 558 | | Differential | | | X | | |
| 623 | | | | Measurement wheel slipping | X | | |
| 812 | | | Pulley and rubber bands | | | X | |
| 1111 | | Multi-resolution Readout | | | X | | |
| 1402 | | Slider | | | | X | |
| 1545 | | Visually count rotations | | | | | X |
| 1553 | | Ticking sound per rotation | | | | X | |
| 1902 | | | | Hard to turn at high loads | X | | |
| 2840 | | | | Doesn't work well on hair | X | | |
| 3049 | | | | Gears not meshing | | | X |
| 3634 | | | | Low/High Gear Reduction | | X | |
| 3786 | | | | Rubber bands don't stay on | | | X |
| 4480 | | | | Measurement is not linear | | | X |
| 4491 | | | | Starting position of the wheel effects measurement | | X | |
| 4509 | | | | Rolling compounds error | X | | |
| 4722 | | | | Double rubber bands around wheel effects measurement | | X | |
| 5050 | | | | Dial mark not visible | | | X |
| 5187 | | | | Calibration is invalid if rubber band slips on pulley | | X | |
| 5370 | | | | A tooth on the dial does not correlate to a rotation | | | X |

Figure 7-2. Spreadsheet summarizing the discoveries design team 3 made during the exercise. Time is in seconds.

### 7.1.8 Design Phase and Process Observations

As proposed in Section 5.2.1.3, the design processes of the teams were observed qualitatively while conducting the experiments and analyzing the audiovisual data.

Special attention was paid to the sequence and duration of the design phases and the timing and the nature of the questions that were asked.

Even though the design phase definitions presented in Section 5.2.1.3, and the conceptual question categories presented in Section 5.2.1.1, provided structure for the observations, for the sake of allowing for the development of a holistic understanding, the activity was not strictly reduced to specific analysis units. Therefore, when investigating the relationship between design process and question asking, the design processes of the teams were not formally "coded," but, rather, evaluated from a broader and more subjective perspective.

Making multiple passes at the data was necessary for gaining that perspective. Overall, each session was observed four times. The initial observation was made during data collection, and was continuos. The second and third observations were made during the identification and analysis of questions and discoveries, respectively, and were composed mainly of discrete and shorter sets of observations, since the nature of the observations required the observer to pause and review different sections of the data. The final observation was continuos, as it was intended to be the final step in ensuring the synthesis of the broader understanding.

## 7.2 Data Analysis and Results

I utilized the data collection and analysis procedures presented in the previous section in performing the analysis in this section. The study of the phenomena outlined in the hypotheses lends itself to three fundamental analysis topics: design performance, question asking and discovery making.

### 7.2.1 Design Performance

I address two aspects of design performance in this section. I first report on the performance of each prototype as measured by the two benchmark metrics, and then cross-validate the metrics by analyzing the performance results they yielded.

### 7.2.1.1 Prototype Performance as Measured by the Benchmark Metrics

I measured the performance of each prototype according to the two external benchmark metrics, by applying the procedures outlined in Section 7.1.4. The results, consisting of the score as measured by M1 and the ranking of each prototype as measured by M2, are shown in Table 7-1. The prototypes which were ranked higher by the experts were assigned a higher number. The ranking assigned by each expert, as well as the average of their rankings, are shown.

| Team | M1 (Score) | M2 (Ave. Rank) | Expert 1 Rank | Expert 2 Rank | Expert 3 Rank |
|---|---|---|---|---|---|
| 1 C | 22 | 4.3 | 2 | 9 | 2 |
| 4 C | 26 | 8.7 | 9 | 7 | 10 |
| 6 C | 11 | 4.7 | 5 | 1 | 8 |
| 8 C | 74 | 12.0 | 12 | 12 | 12 |
| 10 C | 20 | 8.3 | 7 | 11 | 7 |
| 11 C | 49 | 10.7 | 11 | 10 | 11 |
| 2 T | 37 | 6.0 | 10 | 2 | 6 |
| 3 T | 66 | 7.0 | 8 | 8 | 5 |
| 5 T | 31 | 6.0 | 3 | 6 | 9 |
| 7 T | 29 | 4.3 | 6 | 4 | 3 |
| 9 T | 3 | 1.7 | 1 | 3 | 1 |
| 12 T | 22 | 4.3 | 4 | 5 | 4 |

Table 7-1. The performance of each prototype as measured according to the two external benchmark metrics. M1 and M2. The score and the ranking each prototype received are shown. The higher ranked prototypes were assigned a higher number. The ranking assigned by each expert as well as the average of their rankings are shown. The letter C or T. in the team designator. indicates the team belonged to the control or the test group.

### 7.2.1.2 Cross-validating the Benchmark Metrics

Prior to performing analysis regarding the proposed relationships between question asking, discovery making and design performance, it is necessary to cross-validate the benchmark performance metrics M1 and M2. If the metrics cannot be cross-validated, any findings regarding the correlations proposed in H2 and H4 cannot be supported with confidence.

Therefore, I performed correlation analysis between the performance values as measured by M1 and M2 presented in Table 7-1. The result indicates correlation with very high degree of significance (Table 7-2). This finding suggests that the external metrics M1 and M2 are in agreement when they are used to judge the performance of design teams, and constitutes strong evidence for their use as valid benchmark performance metrics when testing for the proposed relationships between question asking, discovery making and design performance.

| | $R^2$ | P |
|---|---|---|
| Judge Ranking vs. Score | 0.55 | **0.006** |

Table 7-2. Correlation coeffient and significance value obtained by performing correlation analysis on the M1 and M2 performance values for each team presented in Table 7-1.

## 7.2.2 Question Asking

In this section, I first present the descriptive statistics for the type of questions that were asked during the twelve design sessions. I then analyze the proposed relationships between question asking and design process, performance, and interaction with hardware. And finally, I take a closer look at the interplay between DRQs and GDQs, and demonstrate the relevance of treating DRQs and GDQs as complementary pairs.

### 7.2.2.1 Descriptive Statistics for the Types of Questions that were Asked

I analyzed the data on the frequency of questions in conjunction with the results of the question categorization process described in 7.1.6 in producing descriptive statistics for the types of questions that were asked during the 12 experiments.

Table 7-3 shows the distribution of the question asking rates among the 22 question categories for each design team. (The results are shown in questions asked per hour.)

**Distribution of Questions among Categories per Team (questions/hr)**

| | Team Designator | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Question Category** | **1 C** | **2 T** | **3 T** | **4 C** | **5 T** | **6 C** | **7 T** | **8 C** | **9 T** | **10 C** | **11 C** | **12 T** |
| Request/Directive | 12.4 | 9.6 | 12.1 | 7.8 | 6.8 | 10.7 | 15.5 | 19.1 | 5.2 | 14.5 | 13.7 | 17.6 |
| Verification | 48.4 | 53.1 | 91.2 | 68.1 | 52.5 | 60.2 | 57.9 | 58.8 | 63.1 | 50.6 | 61.6 | 55.3 |
| Disjunctive | 0.6 | 0.8 | 3.6 | 1.2 | 2.3 | 2.7 | 0.6 | 2.6 | 1.7 | 0.7 | 1.2 | 2.6 |
| Concept Completion | 14.2 | 8.8 | 21.8 | 22.1 | 9.7 | 12.7 | 21.5 | 9.2 | 27.5 | 8.5 | 13.1 | 6.5 |
| Feature Specification | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.6 | 0.0 | 0.6 | 0.7 |
| Quantification | 3.5 | 5.5 | 10.9 | 8.4 | 9.1 | 3.3 | 6.6 | 7.9 | 3.4 | 3.3 | 1.9 | 4.6 |
| Definition | 0.6 | 2.3 | 1.2 | 0.6 | 1.1 | 0.7 | 0.6 | 2.0 | 0.6 | 0.0 | 0.0 | 2.6 |
| Example | 0.0 | 0.0 | 0.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Comparison | 1.2 | 0.0 | 2.4 | 2.4 | 2.3 | 2.0 | 1.2 | 0.0 | 0.6 | 0.0 | 0.0 | 0.0 |
| Judgemental | 8.9 | 4.9 | 8.5 | 10.1 | 8.6 | 10.7 | 4.8 | 10.6 | 4.6 | 9.2 | 1.9 | 7.2 |
| Interpretation (DRQ) | 3.5 | 2.7 | 6.6 | 4.2 | 5.1 | 4.0 | 2.4 | 5.3 | 3.4 | 0.7 | 4.4 | 3.3 |
| Procedural (DRQ) | 1.2 | 1.0 | 0.0 | 0.0 | 0.6 | 2.0 | 1.8 | 0.7 | 1.1 | 0.7 | 0.6 | 1.3 |
| Causal Antecedent (DRQ) | 0.0 | 0.0 | 0.6 | 0.6 | 0.6 | 0.0 | 1.2 | 0.0 | 0.6 | 0.0 | 0.6 | 0.0 |
| Causal Consequence (DRQ) | 0.6 | 0.0 | 0.0 | 0.6 | 0.6 | 0.0 | 0.0 | 0.0 | 0.6 | 0.0 | 0.0 | 0.0 |
| Rationale/Function (DRQ) | 1.8 | 4.5 | 1.8 | 2.4 | 5.7 | 0.0 | 3.0 | 3.3 | 1.1 | 1.3 | 3.1 | 1.3 |
| Expectational (DRQ) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Enablement (DRQ) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Enablement (GDQ) | 0.6 | 0.8 | 1.2 | 0.6 | 0.0 | 0.0 | 0.6 | 0.0 | 0.6 | 2.6 | 2.5 | 0.7 |
| Method Generation (GDQ) | 6.5 | 2.3 | 8.5 | 2.4 | 2.9 | 5.3 | 3.0 | 2.0 | 3.4 | 5.9 | 2.5 | 2.6 |
| Proposal/Negotiation (GDQ) | 11.8 | 13.7 | 19.3 | 14.3 | 7.4 | 4.7 | 16.7 | 22.5 | 8.6 | 13.1 | 14.9 | 9.1 |
| Scenario Creation (GDQ) | 0.0 | 0.0 | 0.0 | 0.6 | 0.0 | 0.7 | 0.6 | 0.0 | 0.0 | 0.7 | 0.6 | 0.0 |
| Ideation (GDQ) | 3.5 | 3.9 | 1.8 | 1.2 | 2.3 | 2.0 | 3.6 | 1.3 | 1.1 | 3.3 | 1.2 | 2.0 |
| Total Questions | 119.9 | 114.9 | 192.1 | 148.1 | 117.5 | 122.3 | 141.6 | 147.2 | 128.6 | 115.0 | 125.1 | 117.2 |
| Total DRQ | 7.7 | 9.2 | 9.1 | 8.4 | 12.6 | 6.7 | 8.4 | 11.2 | 7.5 | 2.6 | 9.3 | 5.9 |
| Total GDQ | 22.5 | 20.8 | 30.8 | 19.1 | 12.6 | 12.7 | 24.5 | 25.8 | 13.8 | 25.6 | 21.8 | 14.3 |
| Total DRQ+GDQ | 30.1 | 30.0 | 39.9 | 27.5 | 25.1 | 19.4 | 32.9 | 37.0 | 21.2 | 28.3 | 31.1 | 20.2 |

Table 7-3. Distribution of the question asking rates among the 22 question categories for each design team. The results are reported in questions asked per hour. The letter C or T. in the team designator. indicates the team belonged to the control or the test group.

Table 7-4 reports the same set of results, however, the distribution of questions among the 22 question categories for each design team are shown as the percentage of the total questions asked.

**Distribution of Questions among Categories per Team (% of total questions)**

| | Team Designator | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Question Category** | **1 C** | **2 T** | **3 T** | **4 C** | **5 T** | **6 C** | **7 T** | **8 C** | **9 T** | **10 C** | **11 C** | **12 T** |
| Request/Directive | 10.3 | 8.3 | 6.3 | 5.2 | 5.8 | 8.7 | 11.0 | 13.0 | 4.0 | 12.6 | 10.9 | 15.0 |
| Verification | 40.4 | 46.2 | 47.5 | 46.0 | 44.7 | 49.2 | 40.9 | 39.9 | 49.1 | 44.0 | 49.3 | 47.2 |
| Disjunctive | 0.5 | 0.7 | 1.9 | 0.8 | 1.9 | 2.2 | 0.4 | 1.8 | 1.3 | 0.6 | 1.0 | 2.2 |
| Concept Completion | 11.8 | 7.7 | 11.3 | 14.9 | 8.3 | 10.4 | 15.2 | 6.3 | 21.4 | 7.4 | 10.4 | 5.6 |
| Feature Specification | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.4 | 0.0 | 0.5 | 0.6 |
| Quantification | 3.0 | 4.8 | 5.7 | 5.6 | 7.8 | 2.7 | 4.6 | 5.4 | 2.7 | 2.9 | 1.5 | 3.9 |
| Definition | 0.5 | 2.0 | 0.6 | 0.4 | 1.0 | 0.5 | 0.4 | 1.3 | 0.4 | 0.0 | 0.0 | 2.2 |
| Example | 0.0 | 0.0 | 0.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Comparison | 1.0 | 0.0 | 1.3 | 1.6 | 1.9 | 1.6 | 0.8 | 0.0 | 0.4 | 0.0 | 0.0 | 0.0 |
| Judgemental | 7.4 | 4.3 | 4.4 | 6.9 | 7.3 | 8.7 | 3.4 | 7.2 | 3.6 | 8.0 | 1.5 | 6.1 |
| Interpretation (DRQ) | 3.0 | 2.4 | 3.5 | 2.8 | 4.4 | 3.3 | 1.7 | 3.6 | 2.7 | 0.6 | 3.5 | 2.8 |
| Procedural (DRQ) | 1.0 | 0.9 | 0.0 | 0.0 | 0.5 | 1.6 | 1.3 | 0.4 | 0.9 | 0.6 | 0.5 | 1.1 |
| Causal Antecedent (DRQ) | 0.0 | 0.0 | 0.3 | 0.4 | 0.5 | 0.0 | 0.8 | 0.0 | 0.4 | 0.0 | 0.5 | 0.0 |
| Causal Consequence (DRQ) | 0.5 | 0.0 | 0.0 | 0.4 | 0.5 | 0.0 | 0.0 | 0.0 | 0.4 | 0.0 | 0.0 | 0.0 |
| Rationale/Function (DRQ) | 1.5 | 3.9 | 0.9 | 1.6 | 4.9 | 0.0 | 2.1 | 2.2 | 0.9 | 1.1 | 2.5 | 1.1 |
| Expectational (DRQ) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Enablement (DRQ) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Enablement (GDQ) | 0.5 | 0.7 | 0.6 | 0.4 | 0.0 | 0.0 | 0.4 | 0.0 | 0.4 | 2.3 | 2.0 | 0.6 |
| Method Generation (GDQ) | 5.4 | 2.0 | 4.4 | 1.6 | 2.4 | 4.4 | 2.1 | 1.3 | 2.7 | 5.1 | 2.0 | 2.2 |
| Proposal/Negotiation (GDQ) | 9.9 | 11.9 | 10.1 | 9.7 | 6.3 | 3.8 | 11.8 | 15.2 | 6.7 | 11.4 | 11.9 | 7.8 |
| Scenario Creation (GDQ) | 0.0 | 0.0 | 0.0 | 0.4 | 0.0 | 0.5 | 0.4 | 0.0 | 0.0 | 0.6 | 0.5 | 0.0 |
| Ideation (GDQ) | 3.0 | 3.4 | 0.9 | 0.8 | 1.9 | 1.6 | 2.5 | 0.9 | 0.9 | 2.9 | 1.0 | 1.7 |
| Total DRQ | 6.4 | 8.0 | 4.7 | 5.6 | 10.7 | 5.5 | 5.9 | 7.6 | 5.8 | 2.3 | 7.5 | 5.0 |
| Total GDQ | 18.7 | 18.1 | 16.0 | 12.9 | 10.7 | 10.4 | 17.3 | 17.5 | 10.7 | 22.3 | 17.4 | 12.2 |
| Total DRQ+GDQ | 25.1 | 26.1 | 20.8 | 18.5 | 21.4 | 15.8 | 23.2 | 25.1 | 16.5 | 24.6 | 24.9 | 17.2 |

Table 7-4. Distribution of the questions among the 22 question categories for each design team as the percentage of the total questions asked. The letter C or T. in the team designator. indicates the team belonged to the control or the test group.

Finally, Table 7-5 reports a subset of the results, where only the averages for the control and test groups are considered.

**Distribution of Questions among Categories for Control and Test Groups (questions/hr and % of total questions)**

| | Rate (q/hr) | | % of Total | |
|---|---|---|---|---|
| Question Category | Control | Test | Control | Test |
| Request/Directive | 13.0 | 11.1 | 10.1 | 8.2 |
| Verification | 57.9 | 62.2 | 44.7 | 46.0 |
| Disjunctive | 1.5 | 1.9 | 1.2 | 1.4 |
| Concept Completion | 13.3 | 16.0 | 10.3 | 11.8 |
| Feature Specification | 0.1 | 0.2 | 0.1 | 0.2 |
| Quantification | 4.7 | 6.7 | 3.6 | 4.9 |
| Definition | 0.6 | 1.4 | 0.5 | 1.0 |
| Example | 0.0 | 0.1 | 0.0 | 0.1 |
| Comparison | 0.9 | 1.1 | 0.7 | 0.8 |
| Judgemental | 8.6 | 6.4 | 6.6 | 4.7 |
| Interpretation (DRQ) | 3.7 | 3.9 | 2.8 | 2.9 |
| Procedural (DRQ) | 0.9 | 1.0 | 0.7 | 0.7 |
| Causal Antecedent (DRQ) | 0.2 | 0.5 | 0.2 | 0.4 |
| Causal Consequence (DRQ) | 0.2 | 0.2 | 0.2 | 0.1 |
| Rationale/Function (DRQ) | 2.0 | 2.9 | 1.5 | 2.2 |
| Expectational (DRQ) | 0.0 | 0.0 | 0.0 | 0.0 |
| Enablement (DRQ) | 0.0 | 0.0 | 0.0 | 0.0 |
| Enablement (GDQ) | 1.1 | 0.6 | 0.8 | 0.5 |
| Method Generation (GDQ) | 4.1 | 3.8 | 3.2 | 2.8 |
| Proposal/Negotiation (GDQ) | 13.6 | 12.5 | 10.5 | 9.2 |
| Scenario Creation (GDQ) | 0.4 | 0.1 | 0.3 | 0.1 |
| Ideation (GDQ) | 2.1 | 2.4 | 1.6 | 1.8 |
| Total Questions | 129.6 | 135.3 | 100.0 | 100.0 |
| Total DRQ | 7.7 | 8.7 | 5.9 | 6.5 |
| Total GDQ | 21.2 | 19.5 | 16.4 | 14.4 |
| Total DRQ+GDQ | 28.9 | 28.2 | 22.3 | 20.8 |

Table 7-5. Distribution of the questions among the 22 question categories for the control and test groups in terms of the rate at which questions were asked and as the percentage of the total questions asked. Only the averages for the control and test groups are considered.

These results show that close to half of the questions that were asked consisted of Verification questions. That is not surprising as Verification questions operate at the lowest cognitive level and are instrumental in establishing rudimentary communication. The other two types of questions that were asked at a significantly higher rate than others were the Proposal/Negotiation and Concept Completion questions. The high occurrence of Concept Completion questions can be interpreted in a similar way to the Verification questions. However, the high occurrence of Proposal/Negotiation questions is significant as they are GDQs. I will discuss this finding in Chapter 8.

When the question asking rates of the control and test groups during the exercise are compared, the results seem strikingly similar. More specifically, there is no statistically significant difference between the averages of the DRQ-GDQ and overall question asking rates of the two groups.

Comparison of the DRQ and total question asking rates obtained from the design exercise with the ones Graesser obtained from tutoring sessions yields the results shown in Table 7-6[25]. GDQ asking rates during tutoring are not reported since Graesser does not explicitly account for them. Also, since Graesser does not make a GDQ distinction, he most likely accounts for the Method Generation category that I account for in the GDQ class in his DRQ class under the Procedure category. Graesser also accounts for the GDQ Enablement category under his DRQ Enablement category. Finally, Graesser does not consider the Interpretation category as a DRQ, whereas I do[26]. The DRQ rates reported in Table 7-6 are adjusted to account for DRQs in the way Graesser does to allow for comparison. However, how Graesser accounts for the other three GDQ classes is not clear. It is possible those types of questions did not even occur in a tutoring context.

|  | Designing C | Designing T | Tutoring |
|---|---|---|---|
| Total Questions | 129.6 | 135.3 | 116.3 |
| Total DRQ | 9.1[27] | 9.2[27] | 19.8 |
| Total DRQ+GDQ | 28.9 | 28.2 | n/a |

Table 7-6. Comparison of the DRQ and total question asking rates I obtained from the design exercise with the ones Graesser obtained from tutoring sessions (in questions asked per hour). The letter C indicates rates for the control group, and the letter T indicatates rates for the test group.

The data in Table 7-6 show that more DRQs were asked during the tutoring session than the design exercise. Since I have not viewed the data from the tutoring sessions, it is difficult for me to account for the difference. Regardless, one explanation would be obtained by considering that the *nature* of the tutoring session might have promoted the

---

[25] Graesser's empirical findings were presented in Section 3.4.

[26] These differences were discussed in detail in Section 3.5.2.

[27] This DRQ asking rate is different from the one shown in Table 7-5 since my designation of DRQ categories are different than Graesser's. What is shown is the adjusted rate so that DRQs are accounted for in the way Graesser does.

asking of more DRQs; the student and tutor pairs were most likely expected to "converge" on the "subject matter," and spent most of their energy doing so. However, in the design exercise, there was no specific subject matter, and the designers spent a significant portion of their energy in generating ideas and in being creative, which resulted in them asking a significant number of GDQs in conjunction with DRQs. I will discuss the notion of treating DRQs and GDQs as complementary pairs in detail in Section 7.2.2.5.

### 7.2.2.2 Question Asking and Design Process

In the next two sections, I will analyze the proposed relationships between question asking and design process by utilizing the two analysis methods I presented in Section 5.2.1.3.

#### 7.2.2.2.1 Question Asking and Design Phase

Monitoring the design processes of the teams and observing if specific question asking rates and question types are associated with each design phase produced valuable insights.

All design teams went through the three fundamental design phases I discussed in Section 5.2.1.3—conceptualization, implementation, and assessment—numerous times during the exercise. As expected, they did so in varying durations, sequences and iterations. Some teams were rather methodical, especially Team 8, and went through them in the above order in general. Other teams, such as Team 6, began by implementation, moved on to conceptualization, back to implementation, and then to assessment. Some teams became predictable, and once they established a phase sequence, they iterated their process by repeating it. Other teams, such as Team 9, were unpredictable, and went in and out of the phases without repeating a pattern. Some teams spent more time in one phase overall than other phases. For instance, Team 5 spent considerably more time than the other teams in the conceptualization phase. Essentially, these observations are a reproduction of the findings of other design researchers.

The significant observation is that such fundamental similarities and differences in the design processes of teams were reflected in the timing and the nature of the questions they asked. Therefore, when monitoring the design process of the design teams, I was able to notice relationships between question asking rates and question types, and design phase. Specially, the most pronounced traits were the teams relying more on GDQs during conceptualization phases than they did during implementation and assessment phases, and more on DRQs during assessment and implementation phases than they did during conceptualization phases (Table 7-7). What I mean by the teams "relying" on a specific class of questions is that class of questions play a comparatively more influential role in their progress, which can be best judged through qualitative evaluation. In many cases, that also meant that they asked a higher number of questions belonging to that class during that phase compared to the number of questions they asked in that class in other phases.

| | **Design Phase** | | |
|---|---|---|---|
| **Question Category** | Conceptualization | Implementation | Assessment |
| Request | ✓ | ✓ | ✓ |
| Verification | ✓ | ✓ | ✓ |
| Disjunctive | ✓ | ✓ | ✓ |
| Concept Completion | ✓ | ✓ | ✓ |
| Feature Specification | ✓ | ✓ | ✓ |
| Quantification | ✓ | ✓ | ✓ |
| Comparison | ✓ | ✓ | ✓ |
| Definition | ✓ | | |
| Judgmental | | | ✓ |
| Interpretation ■ | | | ✓ |
| Procedural ■ | | ✓ | ✓ |
| Causal Antecedent ■ | ✓ | ✓ | ✓ |
| Causal Consequence ■ | | ✓ | ✓ |
| Rationale/Function ■ | | ✓ | ✓ |
| Enablement ● | ✓ | ✓ | |
| Method Generation ● | ✓ | ✓ | |
| Proposal/Negotiation ● | ✓ | ✓ | |
| Scenario Creation ● | ✓ | | |
| Ideation ● | ✓ | | |

Table 7-7. Relationships observed during design activity between question types and design phases. The most pronounced traits were the teams relying more on GDQs during conceptualization phases than they did in implementation and assessment phases, and relying more on DRQs in assessment and implementation phases than they did in conceptualization phases. ■ denotes the types of questions termed as "Deep Reasoning Questions" by Graesser. ● denotes the types of questions termed as "Generative Design Questions" by Eris.

The basis for the qualitative evaluation that lead to the creation of Table 7-7 was the fourth and last pass I made at the data. I started observing each team with an unpopulated version of the matrix presented in Table 7-7 (containing unchecked cells). When I witnessed the asking of a specific type of question having a visible impact on the team's progress, I identified the design phase the team was in, and placed a checkmark in the corresponding box in the matrix. I took "progress" as making a discovery, or gaining critical knowledge and information that might lead to the making of a discovery (a detailed discussion on discovery making is provided in Section 6.1.2). After populating a matrix for each team, I superimposed all of them, and synthesized the general matrix presented in Table 7-7.

In the general matrix, the check marks for each question category represent a relative distribution. For example, if Ideation was checked in six of the team matrixes during Conceptualization, and checked in one or two of the team matrixes during Implementation and Assessment, it was only checked in the general matrix during Conceptualization. Also, three types of questions were not asked at all by any of the teams during the experiments: Example, Expectational and Enablement (DRQ). That is most likely the result of the limited duration of the design exercise. Since I was not able to make any observations on the impact of those types of questions, they are not accounted for in the general matrix.

The associations illustrated in Table 7-7 can be discussed in terms of the principles behind the question categories. Before addressing the distribution of question types to specific phases, I will reflect on the perceived influence of the first seven question categories in all three phases. What I observed in the data suggested that the first seven categories were closely associated with rudimentary communication mechanisms, which were geared toward information exchange and social mediation of the activity. Therefore, it is natural for them to appear to have a similar degree of influence in all three phases; they are too fundamental to be dependent on a specific phase. However, I did notice the teams asking slightly more verification questions in the implementation and assessment phases.

Another type of question that has a similar degree of influence in all three phases is the Causal Antecedent question which aims to uncover the state or events that has caused the question concept. This might point at a fundamental reasoning mechanism that designers utilize in establishing causality. Other reasoning mechanisms, which directly address causality, are embodied in the Casual Consequence and Rationale/Function questions. However, what I observed is that in order for those questions to take on an influential role, concrete events or concepts had to be already constructed. For instance, the Casual Consequence question, "What happened when you pressed it," assumes there was an existing artifact that was operated on. Those types of opportunities for asking Causal Consequence and Rationale/Function questions were less likely to occur in a conceptual phase, where designers—relatively speaking—were not too concerned with firmly grounding themselves in existing events, concepts or artifacts.

When the distribution of question types to specific phases is considered, conceptualization and assessment phases have distinct profiles. Since conceptualization involves tasks geared toward need finding, requirements definition and idea generation, Definition, Scenario Creation and Ideation questions proved to be influential. The other GDQs, Enablement, Method Generation and Proposal/Negotiation questions, were equally influential during conceptualization phases, however they did not contribute to the unique profile as they proved to be pivotal during implementation phases as well.

During assessment phases, Interpretation and Judgmental questions were instrumental in testing the prototype and determining if it met the requirements. In such situations, designers often felt the need to extrapolate the behavior of the prototype they had observed during testing to realistic situations where users would be involved. Interpretation questions played a critical role in extending their observations. Judgmental questions constituted a natural mechanism for initiating and concluding decision making processes.

Implementation phases were rather comprehensive and relied on the asking of a wide range of questions. That was mainly due to the transitional nature of implementation tasks, when designers generated specifications from the needs, requirements, and the concepts which had been defined and generated during conceptualization. Thus, during implementation, the focus was on "generation" as well, but it was more specific and goal driven. Therefore, Procedural, Method Generation, Enablement and Causal Consequence questions were especially influential to team progress.

#### 7.2.2.2.2 Comparison of Meta-Level Understandings

As discussed in Section 7.1.8, I monitored the design processes of the teams by observing each design session four times. During those direct observations, I was able to synthesize a meta-level understanding of how they structured their design tasks, and reflected that structure in their workflow. As an alternative, I gained an understanding of the design process of the teams by considering only the frequency, type and content of the questions they asked. Comparison of those understandings revealed similarities, which complemented and strengthened the results presented in the previous section.

Most teams explicitly considered breaking down their activity into tasks and proposing a structure for their work. As mentioned in the previous section, some were rather methodical, and others seemed to do it just to have the minimum amount of structure they thought was necessary. Teams such as Team 6 did not pay much attention to planning their tasks at all, and improvised as they went along. It can be argued that they did not have structure, and, therefore, their activity should not constitute valid data sources for process observation. If the team did not seem to care for structure, what process was there to study? However, what I observed in their work is that the absence of planned structure resulted in emergent structure of a spontaneous nature, and the resulting activity was worthy of consideration for that reason, if nothing else.

When gaining a meta-level understanding of the design process of each team through direct observation, I paid special attention to a number of descriptive elements of the activity that seemed to be strongly affected by the design processes of the teams. They were:

1) The local goal the team was working toward at any given time.
2) The general topic(s) of discourse. This was usually dependent on the local goal.
3) Change in the direction of discourse. This was usually triggered by the negotiation of the local goal.

4) Social elements such as leadership, and cognitive and political interplay.
5) The level of cognitive progress. This was reflected in the degree of completion of the team's overall design goal.
6) The rate of change in cognitive progress. This was related to the rate at which the team was making conceptual leaps, and, hence, discoveries, and getting closer to accomplishing its overall design goal.

When there was a change in the process of a design team, or, rather, when the team entered a different phase in its design process, that change was usually reflected in the elements I outlined above. More specifically, elements 1, 2, 3 and 4 were reflected in the questions rather strongly and continuously, and elements 5 and 6 were reflected partially and sporadically. By repeatedly observing such effects in the behavior of a team during an entire design session, and by doing so for each of the twelve teams, I was able to form an opinion on the design process of each team.

When gaining a meta-level understanding of the design process of each team through the questions they asked, I reviewed the spreadsheets where the questions were logged. (A sample section of the spreadsheet for Team 12 was illustrated in Figure 7-1.) I read through each spreadsheet a minimum of three times, considering the frequency, type and content of the questions, and attempted to identify and track the descriptive elements listed above. By synthesizing the elements that I was able to identify and track from the spreadsheets, I constructed a second understanding on the design process of each team. I then compared that understanding with the initial, and more accurate, understanding I gained through direct observation.

After performing that analysis for each team, I reached the following conclusion: The *fundamentals* of how a design team structured its design tasks could be reconstructed by gaining exposure to the frequency, type and content of the questions they asked. Even though I consider this to be a valuable finding, I feel the need to mention two significant limitations that are associated with it.

Firstly, the independence of the two understandings I gained of the design process of each team can be questioned, since in order to compare them I needed to gain one before the other. The insight I thought I gained from analyzing the spreadsheets might had already been with me as I might have had acquired it during observing the activity directly. Since evaluation was qualitative and rather subjective, I have no objective way of refuting that claim. However, I made sure that I performed the two methods independently by allowing for a minimum of two weeks between the time I completed the direct observations and began analyzing the spreadsheets.

Secondly, I would like to stress that the understanding of the design processes of the teams I gained by analyzing the spreadsheets was rudimentary, and does not constitute an undiminished replacement for the understanding I gained by observing the activity directly; at best it constitutes a reduced set. However, that is not to say it is not descriptive enough. On the contrary, it would be most appropriate to characterize it as a topographic representation of the design activity, and, hence, as a roadmap to the design thinking and process of the team.

### 7.2.2.3 Question Asking and Performance

Identifying and categorizing the DRQs and GDQs that were asked during the exercises by following the procedures outlined in Section 7.1.6 enabled me to test the proposed relationships between question asking and performance. Prior to focusing on the GDQ-DRQ pairs as suggested in H2 and H3, it is relevant to test for correlation between the overall question asking rates—without making any distinctions between the type of questions—to ensure there is no correlation. If there is, singling out and focusing on the DRQ-GDQ pairs might diminish in meaning.

The combined GDQ+DRQ and overall question asking rates, and the prototype scores for each design team are shown in Table 7-8. Also shown are the averages for the test and control groups.

**Question Asking Rates and Prototypes Scores per Team and Averages for the Control and Test Groups (questions/hr)**

| | 1 C | 2 T | 3 T | 4 C | 5 T | 6 C | 7 T | 8 C | 9 T | 10 C | 11 C | 12 T | C | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Total Questions | 119.9 | 114.9 | 192.1 | 148.1 | 117.5 | 122.3 | 141.6 | 147.2 | 128.6 | 115.0 | 125.1 | 117.2 | 129.6 | 135.3 |
| Total DRQ+GDQ | 30.1 | 30.0 | 39.9 | 27.5 | 25.1 | 19.4 | 32.9 | 37.0 | 21.2 | 28.3 | 31.1 | 20.2 | 28.9 | 28.2 |
| Score | 22.2 | 36.6 | 65.7 | 25.7 | 30.9 | 11.0 | 28.7 | 74.3 | 2.8 | 19.7 | 48.5 | 21.5 | 33.6 | 31.0 |

Table 7-8. The combined GDQ-DRQ and overall question asking rates. and the prototype scores for each design team. The averages for the test and control groups are shown in the last two columns. The results

are reported in questions asked per hour. The letter C or T, in the team designator, indicates the team belonged to the control or the test group.

I already reported that there were no statistically significant differences between the averages of the DRQ+GDQ and overall question asking rates of the two groups in Section 7.2.2.1. Analysis of the prototype score data shown in Table 7-8 yielded the same result for the differences between the averages of the scores of the two groups.

When the overall question asking rates of the twelve design teams were plotted against their prototype scores, no correlation was visible (Figure 7-3). Statistical analysis yielded weak correlation coefficients with low significance, and confirmed this observation (Table 7-9, row 2).

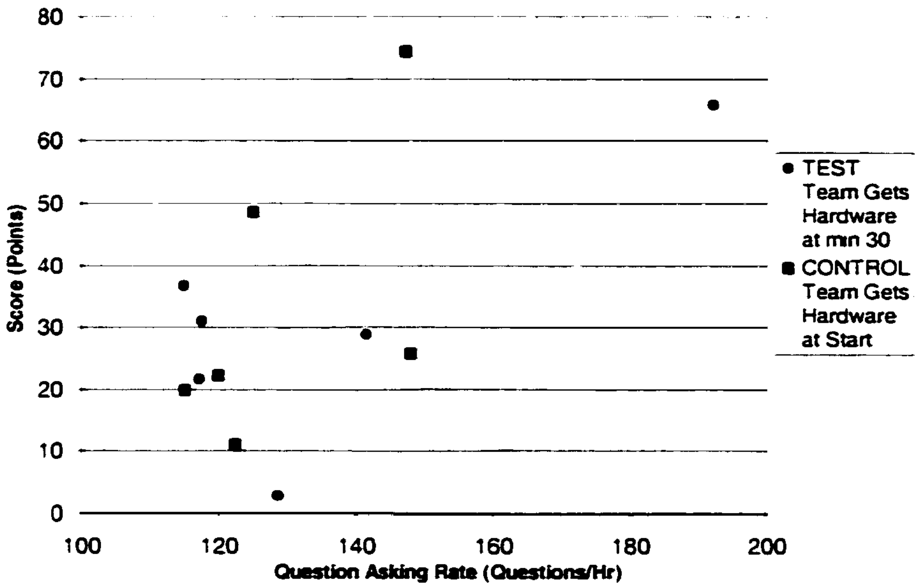


Figure 7-3. Overall question asking rates of the twelve design teams plotted against their prototype scores. Data points marked by squares belong to the teams in the control group, and points marked by circles belong to the teams in the test group.

However, when the combined DRQ+GDQ asking rates of the twelve design teams were plotted against their prototype scores, a linear relationship suggesting positive correlation was visible (Figure 7-4).

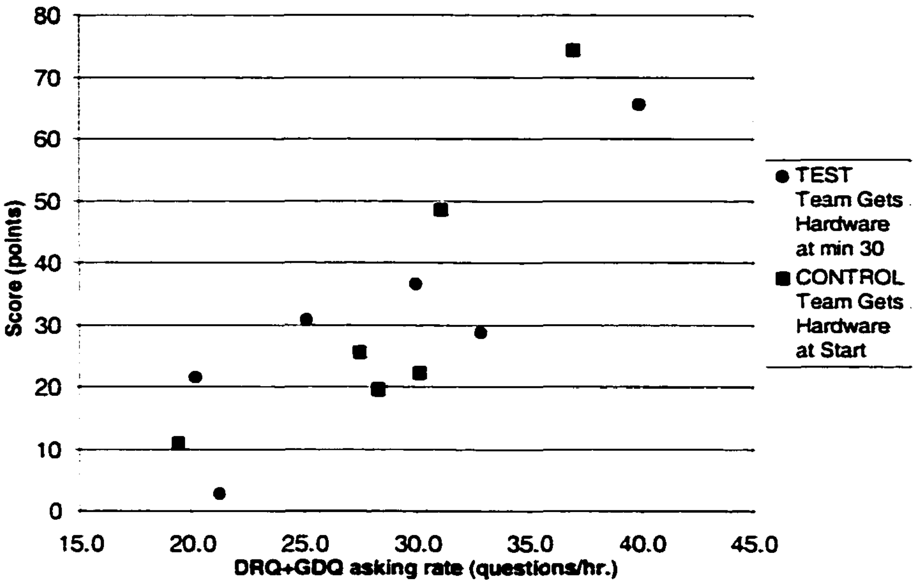


Figure 7-4. Combined DRQ+GDQ asking rates of the twelve design teams plotted against their prototype scores calculated according to M1. Data points marked by squares belong to the teams in the control group. Data points marked by circles belong to the teams in the test group.

Statistical analysis of the data plotted in Figure 7-4 yielded strong correlation coefficients with high significance values (Table 7-9, row 1) for both the control and the test groups.

| | Control $R^2$ | Test $R^2$ | Control P | Test P |
|---|---|---|---|---|
| GDQ+DRQ vs. Score | **0.68** | **0.70** | **0.027** | **0.023** |
| All Questions vs. Score | | | | 0.110 |
| DRQ vs. Score | 0.45 | | 0.087 | - |
| GDQ vs. Score | | 0.56 | | 0.054 |

Table 7-9. Correlation coeffients (adjusted $R^2$) and significance values for correlation between team score and GDQ+DRQ, DRQ, GDQ and overall questions asking rates. Bold highlight indicates strong correlation or high significance. Lighter highlight indicates weaker/no correlation or lower/no significance.

In order to ensure that the occurrence of neither DRQs nor GDQs could establish the positive correlation alone, I analyzed the relationships between DRQ and GDQ asking rates and prototype scores for correlation independently. DRQ asking rates of the control teams correlated positively with prototype scores (Table 7-9, row 3). GDQ asking rates of the test teams correlated with prototype scores (Table 7-9, row 4). However, DRQ asking rates of the test teams, and the GDQ asking rates of the control teams did not correlate with the prototype scores. Also, the strength and significance of the correlation between DRQ asking rates of the control team and the GDQ asking rates of the test teams and prototypes scores was much less than the correlation that existed between the combined DRQ-GDQ asking rates of both groups and prototype scores. These findings suggest that DRQs and GDQs need to be treated as complementary pairs when it comes to establishing their value as a design performance metric.

### 7.2.2.4 Question Asking and Interaction with Hardware

Observing the changes in the combined DRQ+GDQ asking rates of the teams in the test group as they transitioned from the initial part of the experiment, Part A, where they were encouraged to conceptualize in the absence of prototyping hardware to the second part of the experiment, Part B, where they were given access to hardware, and comparing those changes to the changes in the combined DRQ+GDQ asking rates of the teams in the control group during the corresponding time intervals, yielded the results necessary for evaluating H3. In H3, I postulated that DRQ+GDQ asking rates of design teams change when they transition from working in the absence of hardware to working with hardware.

The results are striking as the average of the combined DRQ+GDQ asking rate of the teams in the test group decreased by 21% from Part A to Part B, whereas it increased by 3% for the teams in the control group (Figure 7-5).

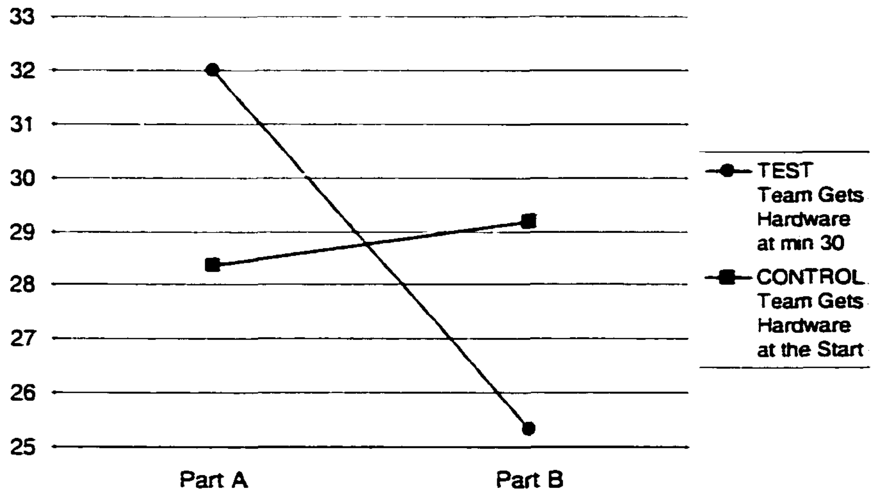


Figure 7-5. Averages of the combined DRQ+GDQ asking rates of the teams in the test and control groups in Parts A and B of the experiment.

The difference between the averages of the combined GDQ+DRQ asking rates for the test group was statistically significant, whereas the difference between the averages for the control group was not (Table 7-10, row 1). Therefore, it can be concluded that the average of the GDQ+DRQ asking rate for the test group decreased significantly, while it did not exhibit any meaningful change for the control group between parts A and B of the experiment.

| | Control P | Test P |
|---|---|---|
| Part A vs. Part B — GDQ+DRQ | [illegible] | 0.063 |
| Part A vs. Part B — GDQ | [illegible] | 0.104 |
| Part A vs. Part B — DRQ | [illegible] | [illegible] |

Table 7-10. Significance values for the difference of the average of the combined GDQ and DRQ, GDQ, and DRQ asking rates of the control and test teams between Part A and Part B of the experiment.

Further analysis showed that the decrease in the average of the combined DRQ+GDQ asking rate of the test teams was directly associated with the decrease in the average of their GDQ asking rate (Figure 7-6), since the averages of their DRQ asking rate did not change significantly (Table 7-10, rows 2 and 3).

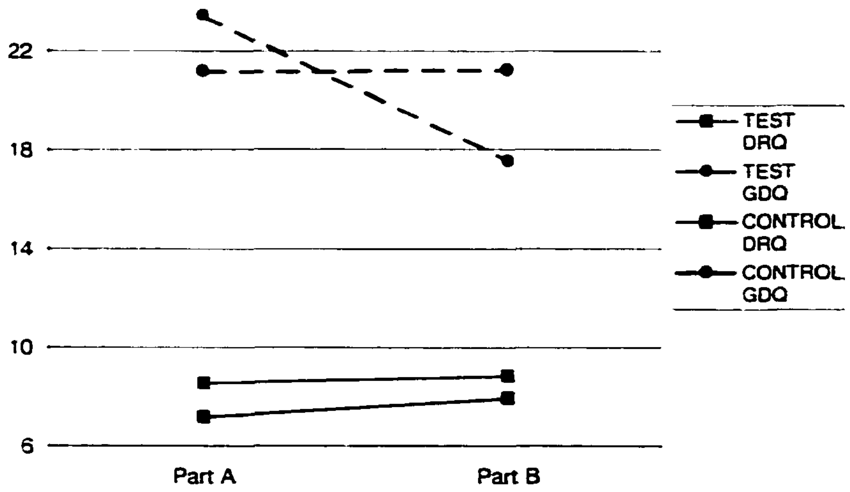


Figure 7-6. Averages of the DRQ and GDQ asking rates of the teams in the test and control groups in Parts A and B of the experiment.

These findings demonstrate that the combined GDQ+DRQ asking rate of the design teams in the test group initially working in the absence of prototyping hardware decreased when they transitioned to working with hardware, and that the combined GDQ+DRQ asking rate of the design teams in the control group did not exhibit any significant change between the corresponding time intervals. These results demonstrate that question asking behavior of design teams is influenced by their access to hardware.

#### 7.2.2.5 DRQs and GDQs as Complementary Pairs

The findings I reported in Section 7.2.2.3 suggest that DRQs and GDQs need to be treated as complementary pairs when it comes to establishing their value as a design performance metric. Based on the data I have collected within the scope of this research, there are at least three additional analysis methods that can be performed in order to gain more insights on that relationship.

The first method would be to hypothesize that there is an optimal DRQ to GDQ asking ratio, and to investigate the relationship between the DRQ/GDQ asking ratios and performance for each team. The second method would be to hypothesize that there are cyclic relationships between DRQs and GDQs, to identify the transitions between DRQs and GDQs, and test for correlation between their DRQ-GDQ transition rates and performance. The third method—the most complex one—would be to hypothesize that there is causality between DRQs and GDQs, and to analyze the data for patterns which might reveal causality links between the occurrences of DRQs and GDQs.

At this stage of the research, time constraints were the driving factor in me choosing to realize the first and second methods only. In applying the first method, I calculated the DRQ/GDQ asking ratios for each team, which are reported in Table 7-11, row 1. When the DRQ/GDQ asking ratios are plotted against the prototype scores for each team, an optimal ratio is not visible. However, it is clear that 10 of the 12 design teams asked approximately 4 DRQs for every 10 GDQs. Even though this observation does not have any significance in suggesting a relationship between DRQ/GDQ asking ratios and performance, it does suggest that 0.4 might be a fundamental DRQ/GDQ ratio in the context of designing.

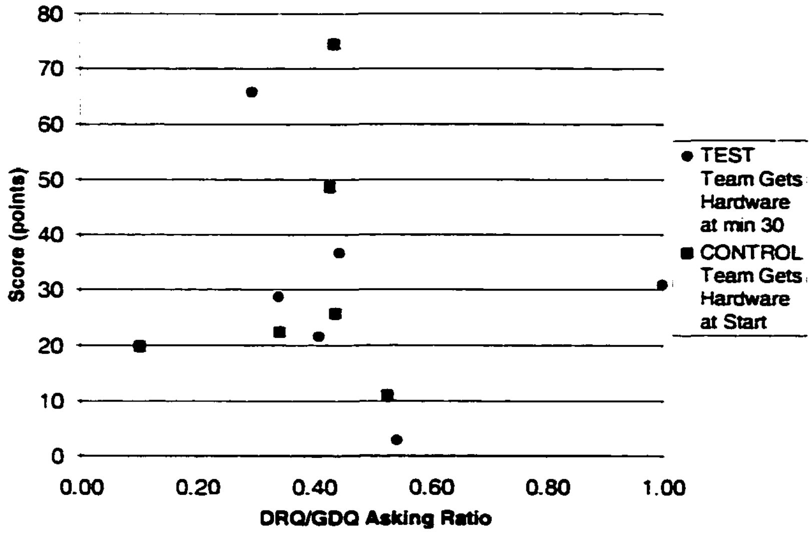


Figure 7-7. DRQ/GDQ asking ratios of the design teams plotted against their prototype scores. Data points marked by squares belong to the teams in the control group. Data points marked by circles belong to the teams in the test group.

In performing the second method, I isolated and considered the data on DRQs and GDQs. I chronologically sorted the DRQs and GDQs each team asked, and accounted for the frequency of the transitions between them. The combined DRQ-GDQ asking rates, the prototype scores, and the DRQ-GDQ transition rates for each design team are shown in Table 7-11. Also shown are the averages for the test and control groups.

**Combined DRQ+GDQ Asking Rates, Scores and DRQ-GDQ Transitions per Team and Averages for the Control and Test Groups (/hr)**

| | 1 C | 2 T | 3 T | 4 C | 5 T | 6 C | 7 T | 8 C | 9 T | 10 C | 11 C | 12 T | C | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DRQ/GDQ Ratio | 0.34 | 0.44 | 0.29 | 0.44 | 1.00 | 0.53 | 0.34 | 0.44 | 0.54 | 0.10 | 0.43 | 0.41 | 0.38 | 0.50 |
| Total DRQ+GDQ | 30.1 | 30.0 | 39.9 | 27.5 | 25.1 | 19.4 | 32.9 | 37.0 | 21.2 | 28.3 | 31.1 | 20.2 | 28.9 | 28.2 |
| Score | 22.2 | 36.6 | 65.7 | 25.7 | 30.9 | 11.0 | 28.7 | 74.3 | 2.8 | 19.7 | 48.5 | 21.5 | 33.6 | 31.0 |
| DRQ-GDQ Transitions | 10.0 | 14.7 | 14.5 | 10.7 | 9.1 | 6.7 | 11.3 | 19.1 | 9.8 | 5.3 | 12.4 | 9.1 | 10.7 | 11.4 |

Table 7-11. The combined GDQ-DRQ asking and DRQ-GDQ transition rates. the prototype scores. and the DRQ/GDQ asking ratios for each design team. The averages for the test and control groups are shown in the last two columns. The results are reported in questions asked and transitions made per hour. The letter C or T. in the team designator. indicates the team belonged to the control or the test group.

Statistical analysis yielded strong correlation of high significance between the DRQ-GDQ transition rates and prototype scores for the control group, but not for the test group (Table 7-12, row 1). The difference between the results of the test and control groups might be related to the behavior illustrated in Figure 7-6—potential natural transition patterns might have been affected by the intervention.

|  | Control $R^2$ | Test $R^2$ | Control P | Test P |
|---|---|---|---|---|
| DRQ-GDQ Transitions vs. Score | **0.85** | 0.41 | **0.005** | 0.101 |
| DRQ+GDQ Asking vs. Transitions | 0.55 | 0.56 | 0.055 | 0.053 |

Table 7-12. Correlation coeffients (adjusted $R^2$) and significance values for correlation between team DRQ-GDQ transition rate and prototype score, and DRQ, DRQ-GDQ transition and combined DRQ+GDQ askings rates. Bold highlight indicates strong correlation or high significance. Lighter highlight indicates weaker/no correlation or lower/no significance.

When interpreting the strong correlation between the DRQ-GDQ transition rates and prototype scores for the control group, it is necessary to keep in mind that the teams that ask more DRQs+GDQs score higher (Table 7-9, row 1). Therefore, it is also necessary to consider that the teams that ask more DRQs+GDQs will be more likely to execute more DRQ-GDQ transitions. Statistical analysis generates supports for that explanation; there is significant correlation between DRQ-GDQ transition and asking rates (Table 7-12, row 2). More analysis is required to determine the extent the relationship between DRQ+GDQ asking rates and the score might be contributing to the correlation between DRQ-GDQ transitions and the score.

Even though the results of the two analysis methods I discussed in this section do not allow me to reach any significant conclusions, they strongly suggest that studying the interplay between the DRQ-GDQ pairs further might prove to be revealing. The third analysis method I mentioned would most likely be instrumental in gaining that understanding.

## 7.2.3 Discovery Making

In this section, I present and categorize the discoveries that were made during the twelve design sessions, and analyze the relationships between discovery making and performance, and discovery making and question asking.

### 7.2.3.1 Categorization and Logging the Discoveries that were Made

I identified the discoveries made by each design team according to the definitions and procedures outlined in sections 7.1.7. After logging the discoveries made by each team in separate spreadsheets as illustrated in Figure 7-2, I merged them into a single spreadheet where all discoveries made by all teams were accounted for under four categories (Figure 7-8).

Overall, 38 discoveries were made regarding the measurement, readout and mechanism concepts, and 31 discoveries were made regarding the obstacles. Qualitative examination of the discoveries reveals that the teams were able to generate ideas that conceptually differ from each other and are rather unique despite the limitations of the laboratory setting. Those findings demonstrate that a wide range of discoveries—quantitatively and conceptually—were made during the experiments, and suggest that the experiment was successful in generating design activity as opposed to problem solving.

| Measurement Concept | Readout Concept | Mechanism | Obstacle | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Rolling a wheel--translating rotation into distance | | | | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| Extrapolate from a standart body part | | | | | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | | | ■ |
| Series of linkages | | | | ■ | | ■ | | | ■ | | ■ | | | | ■ |
| "Set Lengths", a Rod | | | | ■ | ■ | ■ | | ■ | | | ■ | | | | |
| Paper-Pencil outline | | | | ■ | ■ | | | ■ | ■ | | | | | | |
| Interchangable/multiple wheels | | | | | | | | ■ | | | ■ | ■ | ■ | | |
| Rubber Bands String | | | | | ■ | | | | | | ■ | | | ■ | |
| Stationary Device User moves hand | | | | | | | | | | | | ■ | ■ | ■ | |
| Calipers | | | | | | | | | | ■ | | | | | ■ |
| Ratchet | | | | | | | | | | | | | ■ | ■ | |
| Hand Displacement | | | | ■ | | | | | | | | | | | |
| Tank Tracks | | | | | | | | ■ | | | | | | | |
| Separate devices for the measurements | | | | | | | | | ■ | | | | | | |
| Tight-hat mechanism | | | | | | | | | | ■ | | | | | |
| Milling machine | | | | | | | | | | ■ | | | | | |
| Thumb dial | | | | | | | | | | ■ | | | | | |
| Tweezer | | | | | | | | | | | ■ | | | | |
| Rolling a noncircular object | | | | | | | | | | | ■ | | | | |
| Device conforming to hand | | | | | | | | | | | | | | ■ | |
| Pivoting links traversing the contour | | | | | | | | | | | | | | | ■ |
| | Dial | | | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| | Visually count rotations | | | ■ | ■ | ■ | ■ | ■ | | ■ | | ■ | | ■ | |
| | Multi-resolution Readout | | | ■ | ■ | ■ | ■ | | | ■ | | ■ | ■ | ■ | |
| | Ticking sound per rotation | | | ■ | | ■ | ■ | ■ | | | | ■ | ■ | | ■ |
| | Dial rotates twice | | | | | | ■ | | | ■ | ■ | ■ | | ■ | |
| | Flipping Magnet | | | ■ | | | | ■ | | | | | | | |
| | Physical Memory (automatic mark per turn) | | | | ■ | ■ | | | | | | | | | |
| | Slider | | | | ■ | ■ | | | | | | | | | |
| | Physical Memory (manual mark per turn) | | | | ■ | | | | | | ■ | | | | |
| | Differential | | | | | ■ | | | | | | | | | |
| | Winding rubber band unwinds string | | | | | | | ■ | | | | | | | |
| | | Gears | | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| | | Gear/Pulley Reduction | | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | ■ | ■ | |
| | | Rubber band around measurement wheel | | ■ | ■ | | ■ | ■ | | ■ | ■ | ■ | ■ | ■ | |
| | | Pulley and rubber band | | ■ | ■ | ■ | ■ | | | ■ | ■ | | | | |
| | | Wheel rotates arm which ticks the read-out | | | | ■ | ■ | | | | | ■ | ■ | ■ | |
| | | Eccentric cam | | | | | ■ | | | | | | | | |
| | | Rack moving dial | | | | | | | | | | | ■ | | |
| | | | Measurement wheel slipping | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | |
| | | | Not enough gear ratio | | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | ■ | ■ | |
| | | | Low Resolution | ■ | | ■ | ■ | ■ | | | ■ | | | ■ | |
| | | | Device-User interference | ■ | | | ■ | ■ | | ■ | ■ | ■ | | | |
| | | | Measurement piece not fitting between fingers | | ■ | ■ | | | ■ | ■ | ■ | ■ | | | |
| | | | Doesn't work well on hair | | ■ | ■ | ■ | ■ | ■ | | | | ■ | ■ | |
| | | | Negotiating sharp angled countours and corners | | | ■ | | ■ | ■ | | ■ | | | | ■ |
| | | | Too much friction in the drive | | | ■ | ■ | | | | ■ | ■ | | | |
| | | | Wheel shaft doesn't spin well | | | | ■ | ■ | | ■ | | | | ■ | |
| | | | Meshing gears too tight | | | ■ | | ■ | | ■ | | | | | |
| | | | Rolling compounds the error | | | ■ | ■ | | | | | | | | ■ |
| | | | Too much tension on pulley | ■ | | | | | | | | | ■ | | |
| | | | Gears not meshing | ■ | | ■ | | | | | | | | | |
| | | | Rubber bands don't stay on | | ■ | ■ | | | | | | | | | |
| | | | Limited data set for extrapolation | | | | | | ■ | | | | | | ■ |
| | | | Self application is difficult | | | | | | ■ | ■ | | | | | |
| | | | Drive mechanism getting stuck | | | | | | | ■ | ■ | | | | |
| | | | Magnetic force too strong | ■ | | | | | | | | | | | |
| | | | Measurement is nonlinear | | | ■ | | | | | | | | | |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Double rubber bands around wheel effects measurement | | | ■ | | | | | | | | | |
| | | | Dial mark not visible | | | ■ | | | | | | | | | |
| | | | Calibration is invalid if rubber band slips on pulley | | | ■ | | | | | | | | | |
| | | | A tooth on the dial does not correlate to a rotation | | | ■ | | | | | | | | | |
| | | | Backlash for the first inch | | | | ■ | | | | | | | | |
| | | | Need to instruct the user | | | | | | ■ | | | | | | |
| | | | Rubber bands slip | | | | | | | | ■ | | | | |
| | | | Wheel turning too fast for observation | | | | | | | | | ■ | | | |
| | | | Dial gets stuck | | | | | | | | | | ■ | | |
| | | | Dial "jumps"--does not rotate smoothly | | | | | | | | | | | ■ | |
| | | | Does not roll well at an angle | | | | | | | | | | | ■ | |

Figure 7-8. Spreadsheet summarizing all of the discoveries made by all of the design teams. If a team has made a particular discovery, an "X" appears in the cell under the corresponding team column and across the corresponding discovery row. Otherwise, the cell is left blank. In each category, the discoveries that were made by a larger number of teams are listed higher in the table. Darker X's are used for the teams in the control group.

### 7.2.3.2 Discovery Rate and Performance

Identifying the discoveries the design teams made during the experiment allowed me to test H4. The discovery rate, the combined DRQ+GDQ asking rate, and the prototype score for each design team are shown in Table 7-13. Also shown are the averages for the test and control groups.

**Discovery and Combined DRQ+GDQ Asking Rates, and Scores per Team and Averages for the Control and Test Groups (/hr)**

| | 1 C | 2 T | 3 T | 4 C | 5 T | 6 C | 7 T | 8 C | 9 T | 10 C | 11 C | 12 T | C | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Discoveries | 11.8 | 16.4 | 17.5 | 14.3 | 12.6 | 10.0 | 13.1 | 15.8 | 10.9 | 11.2 | 13.1 | 8.5 | 12.7 | 13.2 |
| Score | 22.2 | 36.6 | 65.7 | 25.7 | 30.9 | 11.0 | 28.7 | 74.3 | 2.8 | 19.7 | 48.5 | 21.5 | 33.6 | 31.0 |
| Total DRQ+GDQ | 30.1 | 30.0 | 39.9 | 27.5 | 25.1 | 19.4 | 32.9 | 37.0 | 21.2 | 28.3 | 31.1 | 20.2 | 28.9 | 28.2 |

Table 7-13. The discovery rate, the combined DRQ+GDQ asking rate, and the prototype score for each design team. The averages for the test and control groups are shown in the last two columns. The results are reported in discoveries made and questions asked per hour. The letter C or T, in the team designator, indicates the team belonged to the control or the test group.

When the discovery making rates of the design teams were plotted against their prototype scores, a linear relationship suggesting positive correlation was visible (Figure 7-9).

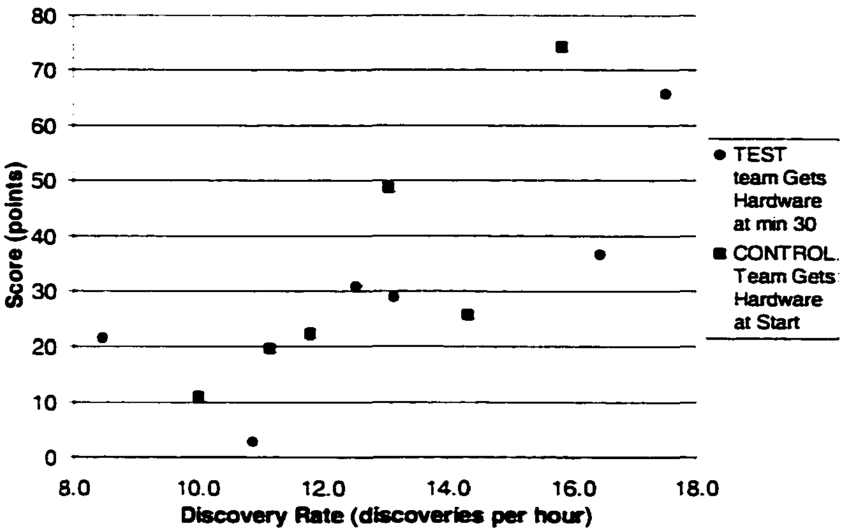


Figure 7-9. Discovery making rates of the twelve design teams plotted against their prototype score. Data points marked by squares belong to the teams in the control group. Data points marked by circles belong to the teams in the test group.

Statistical analysis of the data plotted in Figure 7-9 yielded correlation coefficients with significance for both the control and test teams (Table 7-13). However, the correlation for the test group was not as strong or as significant as the correlation for the control group.

| | Control $R^2$ | Test $R^2$ | Control P | Test P |
|---|---|---|---|---|
| Discovery vs. Score | 0.64 | 0.54 | **0.036** | 0.058 |

Table 7-13. Correlation coeffients (adjusted $R^2$) and significance values for correlation between discovery making rates and prototype scores. Bold highlight indicates strong correlation or high significance. Lighter highlight indicates weaker/no correlation or lower/no significance.

### 7.2.3.3 Discovery Rate and Question Asking

Even though I did not construct a hypothesis regarding the relationship between discovery making and question asking, it was natural to consider that the positive correlation obtained in the previous section between the discovery rates and prototype scores of the design teams might be in part related to the high scoring teams asking more DRQs and GDQs. This is similar to the situation between DRQ-GDQ transitions, prototype scores, and DRQ+GDQ asking rates considered in Section 7.2.2.5.

Statistical analysis of the data on discovery making and DRQ+GDQ asking yielded strong correlation with high significance for the test group, and significant correlation for the control group (Table 7-14).

| | Control $R^2$ | Test $R^2$ | Control P | Test P |
|---|---|---|---|---|
| Discovery vs. DRQ+GDQ | 0.55 | **0.71** | 0.056 | **0.022** |

Table 7-14. Significance values for the difference of the average of the combined GDQ and DRQ, GDQ, and DRQ asking rates of the control and test teams between Part A and Part B of the experiment.

These results suggest that the positive correlation between the discovery rates and prototype scores of the design teams were in part related to the high scoring teams asking more DRQs and GDQs. Therefore, in future research, it would be interesting to analyze the data for patterns which might reveal causality links between the instances of discovery making and occurrences of DRQs and GDQs.

## 7.3 Revisiting the Hypotheses

The analysis results enabled me to evaluate the four hypotheses outlined in Section 6.4. I will now revisit each hypothesis and discuss its validity in light of the findings.

In considering H1, the qualitative analysis presented in Section 7.2.2 demonstrated the following:

1) Specific question asking rates and question types are associated with each design phase.
2) The *fundamentals* of how design teams structure their design tasks can be uncovered by monitoring the frequency, type and content of the questions they ask while designing.

Therefore, focusing on the flow and the nature of the questions asked by design teams can serve as a roadmap to their design thinking, and provides a rudimentary understanding of their design process. This finding validates H1: Question timing and question type are descriptive characteristics of design cognition and process.

The validation of H1 establishes the necessary context for considering H2. In Section 7.1.6, I reported that the trained coders did not experience any difficulties they could not resolve in coding the identified questions according to the 22 categories of the taxonomy of questions, or to the DRQ-GDQ distinction. Those qualitative observations contribute to demonstrating that the principles of the taxonomy of questions, and the DRQ-GDQ distinction are relevant and meaningful.

Also, statistical analysis presented in Section 7.2.3 established strong and significant correlation (adjusted $R^2$ values of 0.68 for the control group and 0.70 for the test group with $p < 0.05$) between combined DRQ+GDQ asking rates of design teams and design performance, whereas significant correlation could not be established between the asking rate of any single type or class of question and design performance. Further analysis presented in Section 7.2.2.5 showed that DRQs and GDQs need to be treated as complementary pairs when it comes to establishing their value as a design performance metric.

When considered in conjunction, those findings validate H2: There exists specific classes of questions, termed Deep Reasoning and Generative Design questions, and their frequency of occurrence within a design team strongly correlates to design team performance, and can be taken as a performance metric.

Testing H3 entailed analyzing the postulated effects of the main intervention of experiment—delaying the introduction of the prototyping hardware to the test teams—on the question asking behavior of design teams. Statistical analysis presented in Section 7.2.2.4 demonstrated that the average of the GDQ+DRQ asking rate for the test group decreased significantly, while it did not exhibit any meaningful change for the control group between parts A and B of the experiment. Further analysis showed that the decrease in the average of the combined DRQ+GDQ asking rate of the test teams was directly associated with the decrease in the average of their GDQ asking rate.

Those findings validate H3: Question asking behavior of design teams is influenced by their access to hardware. DRQ+GDQ asking rates of design teams change when they transition from working in the absence of hardware to working with hardware.

In considering H4, I tested for correlation between the discovery making and DRQ+GDQ asking rates of design teams. The analysis I presented in Section 7.2.3.2 yielded correlation coefficients with significance for both the control and test teams (adjusted $R^2$ values of 0.64 for the control group with $p < 0.10$ and 0.54 for the test group with $p < 0.05$). However, there is a significant limitation to the generalization of this finding.

Since I formulated H4 in a latter stage of this research—while assessing the pilot experiments—the framework I developed in order to characterize and operationalize the phenomena of discovery making had not reached the necessary level of depth by the time the above analysis was conducted in order for me to draw firm conclusions from the results.

Therefore, this finding reiterates the importance of H4, and validates it partially: There is a significant correlation between the frequency of discoveries made by design teams and design team performance. Even though this finding is highly relevant and encouraging, the framework leading to the analysis needs to be developed further and the significance of the correlation needs to be higher ($p < 0.05$) for discovery making to be justified as a performance metric.

# 8 Synthesizing an Understanding of Question Asking while Designing

In the initial three sections of this chapter, I draw upon the findings presented in the previous chapter in developing views for treating question asking as a process, as creative negotiation, and as a mechanism for managing convergent and divergent thinking modes. I then outline the implications of the verified hypotheses in light of those views and list the contributions of this dissertation to design research. In the final section, I discuss opportunities for future research.

## 8.1 Question Asking as a Process

I presented two distinct frameworks in Chapters 3 and 4. The first framework is centered on the questions asked while designing, and characterizes and categorizes them according to their conceptual meaning (Table 3-1). The resulting taxonomy is hierarchical as the lower level question categories are considered to be associated with less sophisticated cognitive mechanisms than the higher level categories. Of particular interest were two classes of questions composed of higher level questions: the Deep Reasoning Questions, which I argued reflect convergent thinking, and Generative Design Questions, which I argued reflect divergent thinking.

The second framework is centered on design performance and conceptualizes it in terms of a series of relationships between four phenomena: design performance, design cognition, design process and question asking (Figure 4-6). The relationships are hierarchical as the lower level phenomena are thought to be a subset of the descriptors of the higher level phenomena. Design process and design cognition are considered to be descriptors of the same level as they are strongly dependent on each other in the sense that one directly feeds, and even causes, the other in a cyclic fashion.

The hierarchical structure of the framework on questions suggests the possibility and relevance of treating question asking as a process. However, since it only articulates the conceptual differences of questions asked while designing, its principles alone are not sufficient in forming a process-centric view of question asking. Even though the hierarchical structure of the taxonomy hints at temporal distinctions, it does not address them explicitly. The element of question asking I investigated solely through empirical means, the timing of questions, provides an initial understanding for the missing temporal dimension required for a process-centric view.

When the principles behind the hierarchical structure of the first framework are considered together with the empirical findings on the timing and nature of questions, it is apparent that treating and investigating question asking *as a process* is feasible and meaningful. Moreover, the characteristic elements of question asking I focused on in this research, the timing and nature of question asked while designing, constituted an explicit link between the two frameworks. Therefore, the hierarchical structure of the second framework provides the means to relate that process-centric view of question asking to the design processes of teams, and ultimately, to their design performance.

In more concrete terms, the rationale presented in the preceding paragraphs is an advanced formulization of what Baya and I have independently observed in the question asking behavior of designers. Baya observes: "The questioning behavior is not random. New questions are being asked after reflecting on information received in answer to a question" [Baya 1996]. The findings of this dissertation lead me to reiterate that initial observation. They also allow me to demonstrate an understanding of the elements of the question asking process by providing specific insights to the nature and timing of questions, i.e. the asking of low level questions forming the necessary knowledge and communication base for the asking of more influential and higher level Deep Reasoning and Generative Design questions.

## 8.2 Question Asking as Creative Negotiation

I reported three significant findings on the use of Proposal/Negotiation questions in design activity in Chapter 7:

1) Approximately ten percent of all of the questions asked by the design teams belonged to the Proposal/Negotiation category (the second most frequently asked question type after the Verification type).
2) Approximately forty percent of all of the Deep Reasoning and the Generative Design questions asked by the design teams belonged to the Proposal/Negotiation category (the design performance metric established in this research is the frequency of occurrence of DRQs and GDQs).
3) The Proposal/Negotiation questions that were asked were most influential during conceptualization and implementation phases of the design process.

These findings demonstrate that Proposal/Negotiation questions play a critical role in the question asking behavior and design performance of teams. However, they do not provide specific insights on the mechanism(s) through which that role is fulfilled. Qualitative consideration of the question asking behavior of the teams during the experiments provided a level of insight and revealed one such mechanism.

Focusing on the temporal dimension of question asking when attempting to uncover that mechanism presented me with a meaningful dilemma: did the concept(s) in the question exist prior to the formulation of the question, or did the formulation of the question lead to its/their creation? These two questions proved to be especially instrumental when I considered them in establishing a context for comparing the temporal dimensions of GDQs with DRQs. Even though one cannot truly resolve that dilemma since the creation of concepts cannot be treated as a discrete phenomenon, and even if it could, there is no objective method of directly measuring what is taking place in someone's mind so that the timing of concept creation could be measured accurately, I will consider it to illustrate the insight I have gained.

The concepts in DRQs exist prior to the formulation of the question. For example, the unknown concept in the Causal Antecedent question: "Why is the wheel spinning?" points at a concept associated with an event that has already taken place—the wheel spinning—and therefore, already exists. Conversely, the concepts in GDQs are created after the formulation of the question. For example, the unknown concept(s) in the Scenario Creation question: "What if the device was used on a child?" points at concept(s) associated with a hypothetical event which has the potential of taking place, and therefore, will be created after the question is formulated. (A detailed discussion on each question category can be found in Section 3.5.2.)

However, Proposal/Negotiation questions constitute an exception; the concept(s) in a Proposal/Negotiation question can already exist, or be created after the formulation of the question as a consequence. More importantly, they can also be created *during* the formulation of the question since most Proposal/Negotiation questions play a transitional role by simultaneously pointing at past and future events or states. That can establish a high degree of conceptual continuity in discourse.

In a team setting, conceptual continuity promotes designers to build on each other's ideas and work more effectively as a group. For example, if the interaction building up to the question: "How about using the wheel instead of the pulley?" is considered, it is very likely that the concept "using a wheel" has occurred to the questioner right before the communication of the question *while* he/she was formulating the question, and that the concept "using a wheel" had been proposed earlier by another person. While asking the question, the questioner creates a spontaneous link between a proposed concept (in the past) and a newly generated hypothetical concept (in the future).

This type of cognitive interplay Proposal/Negotiation questions promote constitutes a mechanism for influencing the design performance of teams, and suggests and supports the notion of treating question asking as "creative negotiation."

## 8.3 Question Asking as a Mechanism for Managing Convergent and Divergent Thinking Modes

The findings reported in Section 7.2.2.1 demonstrated that design teams rely more on GDQs when conceptualizing, and more on DRQs when implementing and assessing in order to make conceptual leaps and advance their designs (Table 7-7).

More specifically, during conceptualization, design teams rely on GDQs by utilizing them as agents of divergent thinking, which entails reframing of previously recognized needs and other existing understandings that establish context for the activity, generation of alternatives, and negotiation (creative reproposal) of proposed design concepts. In general, those events contribute to preserving or increasing ambiguity. The formulation of GDQs in order to initiate convergent thinking modes is not a random event. Rather, it is a conscious effort on behalf of design teams, and can be seen as a response to sensing a need for creativity. Teams continue to rely on the formulation of GDQs and exhibit divergent thinking until that need is satisfied.

During implementation and assessment, design teams rely on DRQs by utilizing them as agents of convergent thinking, which entails focusing on solutions, reiteration and focusing on goals, seeking and establishing causality, and reducing the number of alternatives. In general, those events contribute to reducing ambiguity. As is the case of GDQs, the formulation of DRQs is not a random event either. It can be seen as a response of design teams to sensing a need for being specific and attaining closure. Teams continue to rely on the formulation of DRQs and exhibit convergent thinking until that need is satisfied.

However, this comparison does not imply that design teams simply stop asking DRQs when exhibiting divergent thinking, and stop asking GDQs when exhibiting convergent thinking. As mentioned in Section 7.2.2.1, what I mean by design teams "relying" on a specific class of questions is that class of questions playing a comparatively more influential role in their progress, which can be best judged through qualitative evaluation. In many cases, that also means that design teams ask a higher number of GDQs when conceptualizing compared to the number of GDQs they ask when implementing and assessing, and vice versa, which results in the DRQ/GDQ ratio to change. The findings on DRQ+GDQ asking rates of design teams when working with and without hardware support this view[28]; the DRQ/GDQ ratio increased due to a slight increase in the DRQ asking rates and a significant decrease in GDQ asking rates for the test teams when they transitioned from working in the absence of hardware to working with hardware.

These relationships between GDQ-DRQ usage and divergent-convergent thinking of design teams suggest and support the notion of treating question asking as a mechanism for managing divergent and convergent thinking modes.

## 8.4 Implications of the Verified Hypotheses

When the verified hypothesis is considered in conjunction with the discussion in the previous sections in this chapter, the following conclusions can be drawn:

1) Question asking reflects key aspects of design thinking and processes of teams. Furthermore, design thinking of teams evolves while question asking. While formulating questions—formulation of each question can be considered to be a micro-design task—design teams find the opportunity to structure their design thinking by diverging and converging on their ideas.
2) The frameworks developed in Chapter 3 for characterizing and categorizing questions according to their conceptual meaning, and in Chapter 4 for measuring design performance are valid, and have potential for further development.
3) The question-based metric derived in this study not only measures design performance, but also serves as a descriptive "lens" for revealing and monitoring the thinking of designers during design activity.
4) Question asking, hence design thinking, of teams is strongly influenced by their access to hardware. When conceptualizing in the absence of hardware, design teams exhibit more divergence in their thinking by relying more on Generative Design Questions.

---

[28] Even though the test teams in the experiment went through all three design phases when working with hardware as well as without hardware, they conceptualized more when working without hardware, and implemented and assessed more when working with hardware.

## 8.5 Contributions of The Research

This dissertation makes five contributions to knowledge in the field of design research. It has:

1) Identified a class of questions consisting of five unique question categories which are especially relevant to the questioning processes of designers, termed "Generative Design Questions," as additions to the published taxonomies of questions.
2) Augmented existing methodology for observing design teams in the laboratory by advancing the principles for designing a "design" experiment, and implementing digital audiovisual data collection and analysis techniques.
3) As hypothesized, established the frequency of occurrence of two specific classes of questions as a real-time design performance metric that is internal to design activity.
4) As hypothesized, established the timing and nature of questions as descriptive characteristics of design thinking and process.
5) Demonstrated the feasibility of treating discovery making as a real-time performance metric that is internal to design activity.

## 8.6 Future Research

There are at least five opportunities for future research. I will now briefly discuss each of them and identify principal research questions.

### 8.6.1 Can Asking of more DRQs and GDQs be Promoted?

In the short term, the most pragmatic, and potentially rewarding, research questions to consider are the following: Can a method which promotes the asking of more DRQs and GDQs by design teams be developed, and, if it can, would its application improve design performance of teams?

Answering those questions would entail the development of such a method to promote the question asking processes of design teams, introducing the method to design teams, and then testing their performance. A resulting increase in performance would further validate the findings of this dissertation, and substantiate them by deeming them applicable.

### 8.6.2 Constructing a Framework for Discovery Making in the Context of Question Asking and Design Performance

Another pragmatic research topic that can be addressed in the short term is the detailed investigation of the relationships I have touched upon between DRQ-GDQ asking rates, design performance and discovery making. That would entail constructing an analytical framework that characterizes and operationalizes discovery making while designing, and applying that framework in attempting to identify potential relationships between DRQ-GDQ occurrence sequences, instances of discovery making and design team performance.

### 8.6.3 Real-Time Determination and Display of the Question Asking Metric: An Instrument for Raising Team Performance Awareness

The most fundamental application of the DRQ-GDQ based internal performance metric established in this dissertation is to develop it into an instrument which measures and displays design team performance in real-time. That would be beneficial in providing information on the performance of design teams to themselves as well as to others who share responsibility in their success, such as coaches and managers.

The instrument would be used to increase performance awareness. Design teams can judge their progress according to the reading on the instrument while they are designing. More importantly, support personnel, such as coaches, who have access to limited mechanisms for judging how the design teams they are meant to be supporting are performing other than indirect assessment methods, can utilize the instrument in obtaining a direct measurement and a real-time understanding. That would give them the ability to time and characterize their support, which often comes in the form of constructive interventions, more effectively.

However, the instrument would have limited utility if it were not automated. Real-time automation can possibly be achieved in software by transcribing digitized discourse data, and analyzing the transcripts in order to identify occurrences DRQs and GDQs. However, it is necessary to keep in mind that those are non-trivial tasks, and that they would pose significant challenges.

### 8.6.4 Design Information and Knowledge Systems

Currently, there is a strong interest in the design research community in developing design information and knowledge communication tools. It is almost imperative for such tools to incorporate query based interfaces when accessing and sharing information. The descriptive findings of this research provide a significant part of the necessary understanding for designing such interfaces, and can be transformed into requirements that need to be met if the systems are to support the cognition of designers effectively.

### 8.6.5 Toward a Unified Question-Decision Centric Theory of Design

In the long term, a significant contribution would be to integrate the findings of this dissertation on question asking with existing knowledge on decision making in constructing a design theory. Such an approach can be structured by expanding on the two axiomatic dependencies discussed in Chapter 2 regarding questions and decisions: every question operates on decisions as premises, and conversely, every decision operates on questions as premises.

The implication is that current models of decision making assume the availability of pivotal information when advocating decision making methods without addressing the mechanisms for obtaining the information, and that, if they are viewed in light of the two dependencies, question asking can be taken to be one such mechanism. Developing this approach might result in a new process unifying decision making and question asking, and a new design theory, where question asking attains equal rank as decision making, since high quality questions would yield high quality information, resulting in less ambiguity. In other words, decision making could be viewed as taking place *during* question asking.

# Bibliography

Asimov, M. *Introduction to Design*. Prentice Hall, 1962.

Baya, Vinod, et. al. "An Experimental Study of Design Information Reuse," *Proceedings of the 4th International Conference on Design Theory and Methodology*, ASME, Scottsdale, Arizona, Sept. 13-16, 1992.

Baya, Vinod. *Information Handling Behavior of Engineers in Conceptual Design: Three Experiments*, Ph.D. Dissertation, Stanford University, California, USA, 1996.

Baya, Vinod, Leifer, Larry. "Understanding Information Management in Conceptual Design," *Analyzing Design Activity*. John Wiley & Sons, West Sussex, England, p. 151-167, 1996.

Bloom, S. Benjamin: Editor. *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*, David McKay Company, New York, USA, 1956.

Bloom, S. Benjamin: Editor. *Taxonomy of Educational Objectives, Handbook II: The Affective Domain*, Longman Group, London, England, 1964.

Bucciarelli, L. Louis. "An Ethnographic Perspective on Engineering Design," *Design Studies*, Vol. 9, No. 3, 159-168, July 1988.

Bucciarelli, Louis L. *Designing Engineers*, MIT Press, Cambridge, Massachusetts, USA, 1994.

Brereton, Margot, Cannon, David, Mabogunje, Ade, Leifer, Larry. "Collaboration in Design Teams: How Social Interaction Shapes the Product," *Analyzing Design Activity*, edited by N. Cross, H. Christiaans, K. Dorst, Wiley, 319-341, 1996.

Brereton, Margot. *The Role of Hardware in Learning Engineering Fundamentals: An Empirical Study of Engineering Design and Product Analysis Activity*, Ph.D. Dissertation, Stanford University, California, USA, 1999.

Brereton, Margot. "Distributed Cognition in Design—Negotiating between Abstract and Material Representations," *Proceedings of the 4th International Design Thinking Symposium*, MIT, Massachusetts, USA, 1999.

Carrizossa, Ken, Eris, Ozgur, Mabogunje, Adegboyega, Milne, Andrew, Leifer, Larry. "Building the Design Observatory: a core instrument for design research," *Proceedings of Design 2002*, Dubrovnik, Croatia, 2002.

Cross, Nigel, Christiaans, Henri, Dorst, Kees: Editors. *Analysing Design Activity*, John Wiley & Sons, West Sussex, England, 1996.

Costa, Jorge, et. Al. "An Analysis of Question Asking on Scientific Texts Explaining Natural Phenomena," *Journal of Research in Science Teaching*, V. 37(6), 602-614.

Cuff, Dana. *Negotiating Architecture: A Study of Architects and Clients in Design Practice*, Ph.D. Dissertation, University of California, Berkeley, USA, 1982.

Dieter, E. George. *Engineering Design: A Materials and Process Approach*, McGraw Hill, New York, USA, 1983.

Dillon, T. Jim. "The Classification of Research Questions," *Review of Educational Research*, V. 54, 327-361, 1984.

Dillon, T. Jim. "Questioning in Science," *Questions and Questioning*, ed. By Michel Meyer, New York: De Gruyter, Chapter 4, 68-79, 1988.

Dillon, T. Jim. *Question and Teaching: A Manual of Practice*, Teacher's College Press, New York, USA, 1988.

Dorst, Kees. *Describing Design: A Comparison of Paradigms*, Ph.D. dissertation, Delft University, The Netherlands, 1997.

Eris, Ozgur, Hansen, Poul, Mabogunje, Ade, Leifer, Larry. "Toward a Pragmatic Ontology For Product Development Projects in Small Teams," *Proceedings of the*

*International Conference on Engineering Design*, p. 1645-1650, Munich, Germany, 1999.

Hansen, Poul, Mabogunje, Ade, Ozgur, Eris, Leifer, Larry. "The Product Development Process Ontology: Creating a Learning Research Community," *Proceedings of the International Conference on Engineering Design*, Glasgow, UK, 2001.

Eris, Ozgur, Leifer, Larry. "Facilitating Product Development Knowledge Acquisition: Interaction Between The Expert and The Team," To appear in the International *Journal of Engineering Education special issue on the Social Dimension of Engineering Design*, 2002.

Eodice, Micheal. *A Theory of Requirements Definition in Engineering Design*, Ph.D. Dissertation, Stanford University, California, USA, 2000.

Flammer. A. "Towards a Theory of Question Asking," *Physiological Research*, Vol. 43, 407-420, 1981.

Frankenberger, Eckart, Badke-Schaub, Petra. "Information Management in Engineering Design—Emperical Results from Investigations in Industry," *Proceedings of the International Conference on Engineering Design*, p. 911-916, Munich, Germany, 1999.

Gero, S. John, Editor. *Design Optimization*, Academic Press, Florida, USA, 1985.

Golden, L. James, Jamison, L. David. "Meyer's Theory of Problematology," *Questioning Exchange*, Vol. 2, No. 2, 149-163, 1988.

Graesser, Arthur, Lang, Kathy, Horgan, Dianne. "A Taxonomy for Question Generation," *Questioning Exchange*, Vol. 2, No. 1, 3-15, 1988.

Graesser, Arthur, Golding, Jonathan. "Questioning in Cognitive Psychology and Artificial Intelligence," *Questioning Exchange*, Vol. 2, No. 3, 315-324, 1988.

Graesser, Arthur, Person, Natalie, Huber, John. "Mechanisms that Generate Questions," *Questions and Information Systems*, edited by Thomas W. Lauer, Eileen Peacock, Arthur C. Graesser. Hillsdale, N.J. : L. Erlbaum, Chapter 9, p. 167-187, 1992.

Graesser, Arthur, McMahen, Cathy. "Anomalous Information Triggers Questions When Adults Solve Quantitative Problems and Comprehend Stories," *Journal of Educational Psychology*, Vol. 85, No. 1, 136-151, 1993.

Graesser, Arthur, Person, Natalie. "Question Asking During Tutoring," *American Educational Research Journal*, Vol. 31, No. 1, 104-137, 1994.

Hales, Crispin. *Analysis of the Engineering Design Process in an Industrial Context*, Ph.D. dissertation, Cambridge University, Cambridge, UK, 1987.

Hazelrigg, G. A. "An Axiomatic Framework for Engineering Design," *Journal of Mechanical Design*, Vol 121, p. 342-347, September 1999.

Hazelrigg, G. A. "On The Role and Use of Mathematical Models in Engineering Design," *Journal of Mechanical Design*, Vol 121, 336-341, September 1999.

Howard, Ronald A. "Decision Analysis: Practice and Promise," *Management Science*, v. 34, no. 6, p. 679-695, June, 1988.

Hubka, V. *Principles of Engineering Design*. Translated by W. E. Eder, Butterworth Scientific, Guildford, 1982.

Ijiri, Yuri. *The Foundations of Accounting Measurement*, V. Prentice-Hall Inc., Englewood Cliffs, New Jersey, USA, 1967.

Kerry, T. "Classroom Questions in England," *Questioning Exchange*, Vol. 1, p. 32-33, 1987.

Kuffner, Tom. *Mechanical Design History Content: the Information Requests of Design Engineers*, Master's thesis, Oregon State University, Oregon, USA, 1990.

Kuffner, Tom, Ullman, David. "The Information Requests of Mechanical Engineers," *Design Studies*, Vol. 12, No. 1, 42-50, January 1991.

Lehnert, G. Wendy. *The Process of Question Answering*, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1978.

Lindemann, Udo. "A Model Design Process of Individual Designers," *Proceedings of the 12th International Conference on Engineering Design*, Munich, Germany, 1999.

Mabogunje, Adegboyega. *Measuring Conceptual Design Performance in Mechanical Engineering: A Question Based Approach*, Ph.D. dissertation, Stanford University, California, USA, 1997.

McCracken, James Richard. *Questions: Assessing the Structure of Knowledge and the Use of Information in Design Problem Solving*, Ph.D. Dissertation, Ohio State University, Ohio, USA, 1990.

McGown, A. et. al. "Using Concept Sketches to Track Design Progress," *Proceedings of the 4th International Design Thinking Research Symposium*, MIT, Massachusetts, USA, 1999.

McMahon, C. A., Lowe, A., Culley, S. J. "An Information-Connection Model for Design," *Proceedings of the International Conference on Engineering Design*, p. 1651-1656, Munich, Germany, 1999.

Mendenhall, William, Beaver, Robert. *Introduction to Probability and Statistics*, Ninth Edition, Duxbury Press, California, USA, 1994.

Minneman, Scott. *The Social Construction of a Technical Reality*, Ph.D. dissertation, Stanford University, California, USA, 1991.

Miyake, N. Norman D. A. "To ask a question, one must know enough about what is not known," *Journal of Verbal Learning and Verbal Behavior*, Vol. 18, 357-364, 1979.

Newell, A. and Simon, H. A. *Human Problem Solving*, Prentice-Hall Inc., Englewood Cliffs, New Jersey, USA, 1972.

Pahl, G. "How and Why Collaboration with Cognitive Psychologists Began," Designers—the Key to Successful Product Development, Darmstad Symposium, 1997.

Pahl, G. and Beitz, W. *Engineering Design: A Systematic Approach*, The Design Council, London, England, 1988.

Pugh, Stuart. *Total Design: Integrated Methods for Successful Product Engineering*, Addison-Wesley Publishing Company, Wokingham, England, 1990.

Pugh, Stuart. "Concept Selection—A Method that Works," *Creating Innovative Products using Total Design*, Addison-Wesley, Chapter 14, 167-176, 1996.

Rabinowitz, Mitchell: Editor. *Cognitive Science Foundations of Instruction*, Lawrence Erlbaum Associates, New Jersey, USA, 1993.

Radford, Antony, Gero, S. John, Editor. *Design Optimization*, Gero, S. John, Editor, p.229-258, Academic Press, Florida, USA, 1985.

Shank, R. C. Conceptual Dependency: "A Theory of Natural Language Understanding," *Cognitive Psychology*, 3(4), 552-631, 1972.

Schon, Donald A. *The Reflective Practitioner: How Professionals Think in Action*, Basic Books, New York, USA, 1983.

Schon, Donald A. "Problems, Frames and Perspectives on Designing," *Design Studies*, 5(3), p. 132-136, 1984.

Schon, Donald A. "Teaching and Learning as a Design Transaction," *Research in Design Thinking, Proceedings of Industrial Design Engineering*, p. 21-34, Delft University of Technology, The Netherlands, May 29-31, 21-35, 1991.

Sigel, E. Irving. "Problem Finding in Creativity—A Review of The Creative Vision: A Longitudinal Study of Problem Finding in Art, by J.W. Getzels and M. Csikszentmihalyi," *Questioning Exchange*, Vol. 2, No. 2, 141-147, 1988.

Simon, H. A. *The Science of the Artificial*, 2nd Edition, MIT Press, Cambridge, Massachusetts, USA, 1981.

Tang, C. John. *Toward an Understanding of the Use of Shared Workspaces by Design Teams*, Ph.D. dissertation, Stanford University, California, USA, 1989.

Tang, C. John, Leifer, J. Larry. *An Observational Methodology for Studying Group Design Activity*, Research in Engineering Design, 2, 209-219, 1991.

Ullman, David. *The Mechanical Design Process*, McGraw-Hill, Inc., 1992.

Wilde, D.J. "Using student preferences to guide design team composition," *Proceedings of DETC '97*, Sacramento, 1997.

Willem, A. Raymond. "Design-Science Interactions," *Design Engineering Division publication, ASME*, V. 27, 323-235, 1990.

Willem, A. Raymond. "Design and Science," *Design Studies*, V. 11, No. 1, 42-47, 1990.

Wood, Bill. "A Methodology for Transforming Information into Design Knowledge," *Proceedings of the International Conference on Engineering Design*, p. 131-136, Munich, Germany, 1999.

Yang, Maria. *Retrieval of Informal Information from Design: A Thesaurus Based Approach*, Ph.D. dissertation, Stanford University, California, USA, 2000.

Yen, Samuel. *Capturing Multimodal Design Activities in Support of Information Retrieval and Process Analysis*, Ph.D. dissertation, Stanford University, California, USA, 2000.

# Appendices

## A. Subject Instructions for the Test Group

### Exercise Description/Product Requirements

In this exercise, you will be asked to design and prototype a "bodiometer"; a device that can be moved along the contours of male and female bodies to measure the distance traveled, and hence, the length of body segments—namely, the handweb and the head circumference. The bodiometer must be built from a LEGO parts kit which costs 30 dollars and contains a variety of structural and mechanical components, but no electrical components. No other materials or parts except those supplied with the kit are allowed. Pencil marks may be applied prior to operating the device.

### Performance Criteria

**Handweb** is the perimeter of a hand measured from one side of the wrist to the other, including both sides of the fingers. **Head circumference** is the circumference of the skull measured at eyebrow level.
What drives the overall team score is a combination of sales and cost of your device. The factors that affect sales and cost are explained below.
There will be 7 other design teams carrying out the same exercise. Each team's objective is to maximize their score. Scores will be computed using the following equations:

Score = Sales – Cost

Where

Sales = Design Concept + Aesthetics + Measure Time – Error
Cost = Number of Parts + Manufacturing Time

Variables in these equations are defined as follows:

**Error** is scored as the cumulative absolute value (10 points for 1 inch of error) of the difference between the sum of the two Team measurements and the official measurement where:
Team-measurement = Handweb + Head Circumference
Error = Absolute Value {(team measurement)-(official measurement)}

**Design Concept** is a bonus for a design that provides an instrumented readout, and is worth 50 points. Instrumented readout is any method which allows the user to "read off" a measurement by simply looking at the device without making any calculations or looking at any value tables.

**Aesthetics** is a subjective Bonus category (0-10 points), computed by averaging the scores handed out by a panel of judges (3 design researchers other than the experimenter). Opinions will be based on the device itself. Visual and "intellectual" aesthetics may enter into this opinion.

**Measure Time** is the combined time it takes for the judges to make the two measurements. Points will be handed out in this way: fastest = 20, next fastest = 15, third fastest = 10, 4th fastest = 5, 5th fastest = 3, 6th fastest = 2, 7th fastest=1, and slowest =0.

**Number of Parts** is the total number of parts used in your design. Points will be handed out in this way: highest = 20, 2nd highest = 15, 3rd highest = 10, 4th highest = 5, 5th highest = 3, 6th highest = 2, 7th highest =1, and lowest =0.

**Manufacturing Time** is the time it takes you to rebuild your prototype from an identical and new parts kit after the main part of the experiment is over. . Points will be handed out in this way: highest = 20, 2nd highest = 15, 3rd highest = 10, 4th highest = 5, 5th highest = 3, 6th highest = 2, 7th highest =1, and lowest =0.

## SUGGESTED Schedule and Process

### Phase I--90 minutes

(time)
0:00-0:10 Teams receive the Project Requirements and Performance Criteria worksheet and are encouraged to ask for clarification.

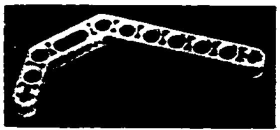
0:10-1:30 *Concept Generation and Prototyping:* The purpose of Phase-I is to explore the design requirements, generate design concepts, and prototype one way of meeting the Product Requirements. The LEGO kit will be provided to you at the beginning of this phase. The deliverable is a functional physical prototype.

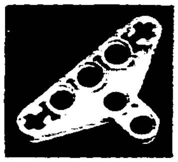
### Phase II--5 minutes

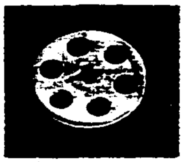
1:30-1:35 *Manufacturing:* In this phase, you will be asked to build a replica of your prototype from an identical and new LEGO parts kit. You may use your existing prototype from Phase II as a reference. The time it takes you to build the replica will be measured and taken as an indicator for the manufacturing time of your design.

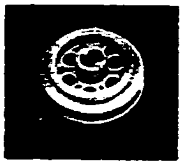
**have FUN!**

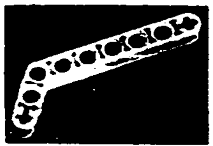
# B. Prototyping Hardware Catalog for the Test Teams

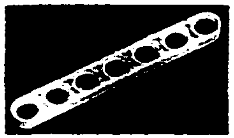

[illegible]

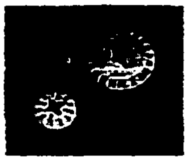

[illegible]

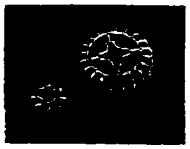

[illegible]

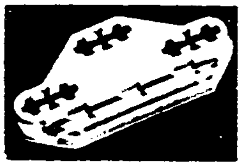

[illegible]

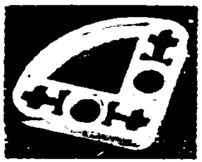

[illegible]

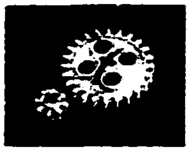

[illegible]

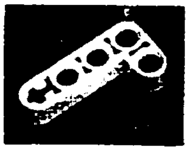

[illegible]

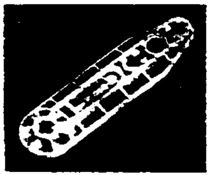

[illegible]

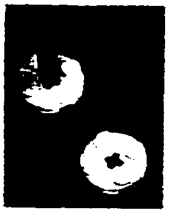

[illegible]

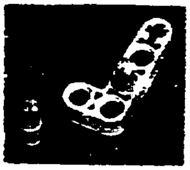

[illegible]

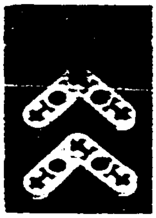

[illegible]

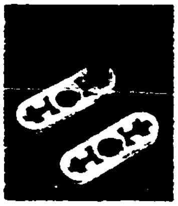

[illegible]

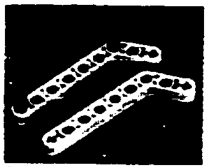


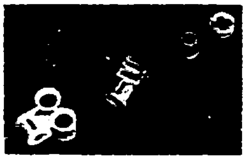




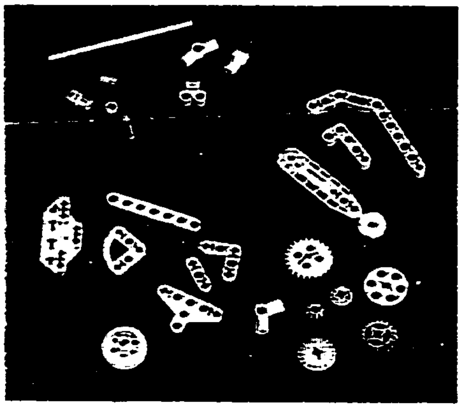









LEGO
TECHNIC
STA
WAR
EPISOD
8001
9+



# C.Sample Transcript (Design Team 1)

| Time In | Time Out | Voice | Utterance |
|---|---|---|---|
| 0:00:00 | 0:00:40 | E | Now this is the real thing. Here's the instructions. It's two pages long. There's something on the back, too. So what I'll let you do is just let you read through it once. And during the exercise, I'll be right outside in this other room. So if you have any questions you can come and just get me. If you knock on this door I'll just come back into the room and we can ask the questions. But I'll just be here for five minutes just to make sure, once you read it, everything's clear. You can still ask questions later but, you know, I'll just be here for five to ten minutes. The schedule's on the back, but you should just kind of read through it, the way it is. |
| 0:01:27 | 0:01:31 | A | YY |
| 0:01:31 | 0:01:38 | B | Wrist. Maybe it's here, besides your fingers. |
| 0:01:38 | 0:01:41 | C | I wonder if it's this way? |
| 0:01:41 | 0:01:49 | B | Ask for it. Oh. Okay. Alright. Fingers. (pause) So it has to be really small. |
| 0:02:11 | 0:02:14 | A | YY |
| 0:02:41 | 0:02:43 | B | Are we actually trying to make this thing? |
| 0:02:43 | 0:02:46 | E | Yeah. You will, yeah you will prototype it with the Lego kit. Yeah. |
| 0:02:46 | 0:02:47 | B | Okay. |
| | | | (E brings in Lego kit) |
| 0:03:40 | 0:03:42 | B | Okay. Star Wars. |
| 0:03:48 | 0:03:49 | C | Do we get to keep this? |
| 0:03.49 | 0:04:08 | E | Yeah. Yeah sure. (pause) So just for your information, I'm running this experiment in two ways. So other groups, you know there's two batches of groups and one group will do it one way and one group will do it another and then I'll compare the two. But I can't tell you before the experiment how they're different- |
| 0:04:08 | 0:04:09 | B | -Sure thing.- |
| 0:04:09 | 0:04:11 | E | -But; you know, so you might not be able to sense what's different. |
| 0:04:11 | 0:04:12 | B | Okay. |
| 0:04:12 | 0:04:13 | E | I'm just letting you know. |
| 0:04:25 | 0:04:32 | E | But both groups are evaluated based on the same, both types of groups will be evaluated based on the same point scheme. |
| 0:06:07 | 0:06:23 | B | I think the problem's going to be around the hand because you're limited by the space. If that can measure the hand accurately then we'll do okay measuring the skull [...] |
| 0:06:33 | 0:06:35 | C | What are we going to try to do? Maximize XX? |
| 0:06:35 | 0:06:36 | B | Yeah. |
| 0:06:36 | 0:06:37 | C | Minimize XX? |
| 0:06:37 | 0:06:38 | B | Yeah. |
| 0:06:38 | 0:06:44 | C | Do you know which, shall we try to concentrate on one of these or try to XX? |
| 0:06:44 | 0:06:46 | B | We should just brainstorm pulling out concepts. |
| 0:06:46 | 0:06:47 | C | Yeah. |
| 0:06:53 | 0:06:55 | B | So, are we, can we start anytime? |
| 0:06:55 | 0:06:56 | E | Yeah. Sure. |
| 0:06:56 | 0:06:59 | B | Okay. Let's look at the parts we have. |
| 0:07:01 | 0:07:04 | A | We could brainstorm without the parts. |
| 0:07:04 | 0:07:06 | C | Yeah. I think that's the best way. |
| 0:07:07 | 0:07:09 | A | So we, we're not limited by them. |
| 0:07:09 | 0:07:12 | B | Alright. Cool. Let's do that. |
| 0:07:28 | 0:07:19 | A | Can we use the board? |
| 0:07:19 | 0:07:25 | E | Yeah, you can use the board. It's on the camera. I can also bring you a sketch pad. I'll go get it. |
| 0:07:27 | 0:07:33 | B | Go ahead. Yeah. Why don't you give YY of designs and I'll go- |
| 0:07:33 | 0:07:37 | C | -Let's just talk about how we want this thing to look like. Like what it's features are going to be. |
| 0:07:40 | 0:07:41 | B | Well, ideally- |
| 0:07:41 | 0:07:43 | C | -Sort of a wheel. Right? |
| 0:07:43 | 0:07:45 | B | Yeah. Ideally I want a wheel. |
| 0:07:45 | 0:07:46 | C | It must have a wheel. |
| 0:07:46 | 0:07:47 | A | Why? |

| | | | |
|---|---|---|---|
| 0:07:47 | 0:07:48 | C | To measure with. |
| 0:07:48 | 0:07:52 | B | Not necessarily because if we don't have that part, if we don't have a round part. |
| 0:07:52 | 0:07:59 | C | Wait is it something that's going to be able to move by itself or are we going to actually move it? |
| 0:07:59 | 0:08:00 | B | We are going to move it. |
| 0:08:00 | 0:08:01 | A | We are going to move it. |
| 0:08:01 | 0:08:02 | B | There is no electrical parts. |
| 0:08:02 | 0:08:03 | A | Yeah. |
| 0:08:04 | 0:08:09 | B | Yeah, I was thinking it would be like a very small container with the wheel--with some sort of-- |
| 0:08:09 | 0:08:10 | C | --we'll be counting-- |
| 0:08:09 | 0:08:10 | B | --rubber- |
| 0:08:10 | 0:08:12 | C | -how many times it goes around- |
| 0:08:12 | 0:08:13 | B | -edge. Yeah exactly.- |
| 0:08:13 | 0:08:14 | C | -and calculate the circumference. |
| 0:08:14 | 0:08:15 | B | Exactly |
| 0:08:16 | 0:08:38 | A | The number of turns. (pause) I was thinking more of something like a string. Okay. Just brainstorming. I don't know how we'd do it with Lego's. You could put a string around it and then stretch it and measure it. That's going to tell you how much it- |
| 0:08:38 | 0:08:44 | C | -And then, how accurate is it going to be? It's not going to like stick to the hand. |
| 0:08:44 | 0:08:45 | A | That's true. |
| 0:08:45 | 0:08:46 | B | Are we allowed to use- |
| 0:08:46 | 0:08:47 | E | -Yeah. You can use the tape measure. |
| 0:08:47 | 0:08:49 | B | -use a tape? So for a measurement? |
| 0:08:49 | 0:09:00 | E | Yeah. And the string if you want to measure it, your head or whatever, perimeter. That's how the official measurements are going to made. By using a string and tape measure. |
| 0:09:00 | 0:09:07 | A | Are we going to be able to use this for, in combination with the Lego, what? |
| 0:09:07 | 0:09:08 | E | No. No.- |
| 0:09:08 | 0:09:09 | B | - It's just the Lego parts. |
| 0:09:09 | 0:09:12 | E | You need to use those parts. Yeah. Nothing outside of those parts. |
| 0:09:12 | 0:09:15 | C | Okay, so, we're basically using that to make it. Just a Lego? |
| 0:09:15 | 0:09:16 | B | Yeah. |
| 0:09:16 | 0:09:17 | E | Yes. That's right. |
| 0:09:17 | 0:09:18 | B | Yeah. So I don't know if we should- |
| 0:09:18 | 0:09:19 | A | -So we can't even do that.- |
| 0:09:19 | 0:09:22 | B | -yeah. I don't know if we should look at it. The parts. |
| 0:09:22 | 0:09:23 | A | Yeah. |
| 0:09:23 | 0:09:30 | B | Because we're totally limited by the parts. [spreading out Lego's] Well. We got a wheel. |
| 0:09:30 | 0:09:31 | A | That's too big. |
| 0:09:31 | 0:09:32 | B | We got. |
| 0:09:33 | 0:09:34 | E | I'll be right outside. |
| 0:09:34 | 0:09:35 | B | Okay.- |
| 0:09:34 | 0:09:35 | C | -Alright- |
| 0:09:35 | 0:09:36 | B | -Thanks. |
| 0:09:42 | 0:09:44 | A | YY |
| 0:09:44 | 0:09:45 | B | Yeah. |
| 0:09:46 | 0:09:49 | C | Are we? Are we being recorded? |
| 0:09:49 | 0:09:50 | A | Yup. |
| 0:09:57 | 0:09:59 | A | We could also [...] |
| 0:09:59 | 0:10:00 | B | Umm. |
| 0:10:00 | 0:10:01 | A | So- |
| 0:10:01 | 0:10:02 | C | -Something that- |
| 0:10:02 | 0:10:03 | A | -wheels- |
| 0:10:03 | 0:10:11 | C | -that counts how many turns. Cause if the wheel's too small, are we going to be able to, like, read it off with our eyes? |
| 0:10:11 | 0:10:12 | A | Right. |
| 0:10:13 | 0:10:15 | C | Well I guess that's, that's what we have to do. |
| 0:10:15 | 0:10:16 | B | We don't have anything XX. |
| 0:10:16 | 0:10:17 | A | Right. |
| 0:10:24 | 0:10:31 | B | Yeah. Ideally, I mean, it would be nice if there was, like a detente, which clicks, like with every, every revolution. |

| | | | |
|---|---|---|---|
| 0:10:31 | 0:10:32 | A | Yeah. |
| 0:10:31 | 0:10:32 | B | Right? |
| 0:10:32 | 0:10:34 | A | See here. We're allowed to make to make a mark- |
| 0:10:34 | 0:10:36 | B | -a mark. Yeah. Let's make a central mark. |
| 0:10:49 | 0:10:50 | B | We have gears. |
| 0:10:50 | 0:11:02 | C | Whereas the design concept, is it bonus for a design that provides an instrumental readout? [reading] Instrumental readout is any method which allows a user to read off a measurement while simply looking at the device- |
| 0:11:02 | 0:11:02 | B | -Right.- |
| 0:11:02 | 0:11:06 | C | -without making a calculation or looking at any value tables. |
| 0:11:11 | 0:11:13 | B | We don't have a lot of good parts here. |
| 0:11:20 | 0:11:22 | B | Do you want to open this? |
| 0:11:22 | 0:11:24 | C | Let's open it here. |
| 0:11:27 | 0:11:28 | A | Not YY |
| 0:11:31 | 0:11:38 | B | Okay. There is a rubber seal. (pause) Rubber seals aren't good because [...] |
| 0:11:47 | 0:11:49 | A | There are (pause) of black things. |
| 0:11:50 | 0:11:52 | B | Oh. It's like a belt. |
| 0:12:05 | 0:12:07 | C | Should we just make this? [looking at Lego plans] |
| | | (laughter) | |
| 0:12:10 | 0:12:11 | B | Yeah. You should. |
| 0:12:14 | 0:12:21 | A | So basically we (pause) want to do this. Right? |
| 0:12:21 | 0:12:22 | C | Yeah. |
| 0:12:22 | 0:12:24 | B | Yeah. Well that's one concept. We shouldn't- |
| 0:12:24 | 0:12:24 | C | -That's one concept- |
| 0:12:24 | 0:12:36 | B | -Yeah. We shouldn't narrow ourselves down to just that. We should keep thinking what else we could measure. How else we could measure our hand. Because this is going to be the bottle neck. Right? |
| 0:12:36 | 0:12:36 | A | Right. |
| 0:12:36 | 0:12:37 | C | This is, like much harder than the skull- |
| 0:12:37 | 0:12:38 | C | -Yeah cause it's- |
| 0:12:38 | 0:12:54 | B | -Yeah. We're limited by space. (pause) Okay. We can have either the wheel. We can have a string, which is clearly not possible with this, these parts. |
| 0:12:57 | 0:12:58 | A | What's that? |
| 0:12:58 | 0:13:03 | B | Like with your string concept? You were saying that we could have a piece of string that runs around- |
| 0:13:03 | 0:13:04 | A | -Yeah but we can't- |
| 0:13:04 | 0:13:05 | B | -yeah but we can't- |
| 0:13:05 | 0:13:06 | C | -we can't use that one- |
| 0:13:05 | 0:13:06 | B | -we can't use that- |
| 0:13:06 | 0:13:07 | A | -No.- |
| 0:13:07 | 0:13:29 | B | -So what else can we do? Other than a wheel? (long pause) Well it sounds really stupid, but what about one bar that floats? Small [...] |
| 0:13:29 | 0:13:33 | A | YY it's almost like a string. |
| 0:13:33 | 0:14:05 | B | Right. (long pause) Right. A fully articulated (pause). Yeah. Basically, a mechanism which has many, many joints in very small sections. Then it is like a snake. Almost. And you can bend it around whatever profile you want. |
| 0:14:07 | 0:14:09 | C | It's going to be really small parts, though. |
| 0:14:09 | 0:14:11 | B | Yeah. It has to be really small. |
| 0:14:11 | 0:14:14 | C | Yeah. Because if you have things like this- |
| 0:14:14 | 0:14:14 | B | -Yeah- |
| 0:14:14 | 0:14:15 | C | -YY- |
| 0:14:15 | 0:14:25 | B | -it won't even. It won't even fit into your hand. Yeah. It has to be like little sections. Like these. Many of them. And they- |
| 0:14:25 | 0:14:29 | A | -The drawback would be that it, it's going to have a lot of parts. |
| 0:14:29 | 0:14:48 | B | Right. And also you won't be able to tell the measurement just by looking at it. Because you. Like after you bend it around your hand you would have to, probably like, mark it. Yeah. You would have to count, either count number of segments or you mark it, stretch it, and measure it. |
| 0:14:48 | 0:15:02 | C | Or you could use the same, same length parts. Then we know how much, how long one is. Like after seeing how many, how many joints we have, we have the links right away. |
| 0:15:02 | 0:15:13 | B | Right. Any other concepts? We want concepts. Concepts. We have gears. |
| 0:15:13 | 0:15:25 | A | Maybe, maybe we can make some assumptions about, like width of the fingers that we can't |

| | | | |
|---|---|---|---|
| | | | reach. YY although my finger's narrower than the cable. |
| 0:15:25 | 0:15:28 | B | Yeah. And then what would you do afterwards? You'd [...] |
| 0:15:28 | 0:15:53 | A | I don't know. I'm just saying shit. You might not have to measure this one here. I mean. Yeah. Otherwise, then it's going to be much simpler because this is straight, here. This is virtually straight. Straight. Straight. So we don't have to have so many joints. |
| 0:15:53 | 0:15:54 | C | Okay. Alright. |
| 0:15:54 | 0:16:02 | B | But then how do you measure (pause) your hand length at one time? Yeah. |
| | | (laughter) | |
| 0:16:04 | 0:16:05 | C | Alright [Subject 1]. |
| 0:16:08 | 0:16:09 | A | I'm just brainstorming. |
| 0:16:09 | 0:16:11 | B | Yeah. Yeah. I know. That's good. |
| 0:16:41 | 0:16:42 | B | Okay. |
| 0:16:56 | 0:16:58 | B | Aarrggh. Man. Come on. |
| 0:17:07 | 0:17:08 | A | Here [...] |
| | | (laughter) | |
| 0:17:13 | 0:17:24 | A | YY volume. Then you can. For example, you have a (pause) some kind of container, filled with water.- |
| 0:17:24 | 0:17:25 | B | -Umm Hummm- |
| 0:17:25 | 0:17:30 | A | -and measure the volume by displacement. If you fill it up with water- |
| 0:17:30 | 0:17:31 | B | -Right.- |
| 0:17:31 | 0:17:33 | A | -and put your hand in there and then water- |
| 0:17:33 | 0:17:34 | B | -Yeah- |
| 0:17:34 | 0:17:35 | A | -flows out. |
| 0:17:35 | 0:17:36 | B | Yeah. That's following. |
| 0:17:36 | 0:17:38 | A | Yeah. Is there something similar- |
| 0:17:38 | 0:17:38 | B | -dimensions- |
| 0:17:38 | 0:17:45 | A | -we can do for length? Can we think of a way [...] |
| 0:17:45 | 0:17:52 | B | Yeah. Just stare at your hand and say, 'Hey that's six point seven-five inches.' |
| 0:17:54 | 0:17:55 | A | Small |
| 0:18:02 | 0:18:10 | B | We pretty much know what parts are available to us, now. Right? I don't know if we should stop staring at them and just think of concepts? |
| 0:18:10 | 0:18:11 | A | Yeah. |
| 0:18:13 | 0:18:14 | B | What do you think? |
| 0:18:15 | 0:18:16 | A | Having fun? |
| 0:18:16 | 0:18:18 | C | It doesn't go in here. |
| 0:18:18 | 0:18:19 | B | Yeah. I know. That's too big. |
| 0:18:19 | 8:28:29 | A | That's big. |
| 0:18:21 | 0:18:22 | B | Great. |
| 0:18:24 | 0:18:26 | A | Can we get a small one? |
| 0:18:26 | 0:18:30 | C | So are those, are those the only concepts that we're going to think about? |
| | | B | No. We want to have more. |
| 0:18:31 | 0:18:32 | C | We want to have more. Right? |
| | | B | Yeah. We want to have more. Let's stop staring--at these parts-- |
| | | C | --yeah let's--stop staring about this. |
| | | B | We pretty much know what's available. |
| | | C | Alright. |
| 0:18:42 | 0:18:43 | B | Do you have any other concepts? |
| | | A | Okay. So let's draw these things- |
| | | B | -So we could have- |
| 0:18:56 | 0:18:58 | C | -Why don't we do some function analysis, man?- |
| | | B | -like a wheel- |
| 0:19:00 | 0:19:01 | C | -structure function.- |
| 0:19:01 | 0:19:11 | B | -which turns and counts. Right? Somehow. Or we can have many, many sections of things- |
| | | C | Yeah. |
| 0:19:13 | 0:19:30 | B | -that bends around Or we can have (pause). What was yours [Subject 1]. The one, that doesn't join up with one another? |
| | | A | Oh. I was just saying for this one we don't have to have such small segments. Or we can avoid having small segments if we just avoided measuring anything else. |