Cognitive Science 30 (2006) 65–93

The Role of Functionality in the Mental Representations of Engineering Students: Some Differences in the Early Stages of Expertise

Jarrod Moss[a], Kenneth Kotovsky[a], Jonathan Cagan[b]

[a]*Department of Psychology, Carnegie Mellon University*
[b]*Department of Mechanical Engineering, Carnegie Mellon University*

Received 8 June 2004; received in revised form 4 May 2005; accepted 27 June 2005

Abstract

As engineers gain experience and become experts in their domain, the structure and content of their knowledge changes. Two studies are presented that examine differences in knowledge representation among freshman and senior engineering students. The first study examines recall of mechanical devices and chunking of components, and the second examines whether seniors represent devices in a more abstract functional manner than do freshmen. The most prominent differences between these 2 groups involve their representation of the functioning of groups of electromechanical components and how these groups of components interact to produce device behavior. Seniors are better able to construct coherent representations of devices by focusing on the function of sets of components in the device. The findings from these studies highlight some ways in which the structure and content of mental representations of design knowledge differ during the early stages of expertise acquisition.

Keywords: Psychology; Expertise; Representation; Engineering

1. Introduction

The knowledge structures and processes that are created during the acquisition of expertise are determined by both general cognitive constraints and the content of the domain in which expertise is being acquired. This means that the content and to some extent the structure of the knowledge representations are determined by the particular domain of expertise. The study of a particular domain of expertise provides information about the specific mental representations and cognitive processes employed by experts in that domain, and this information is potentially beneficial in the design of cognitive aids that can assist experts and in improving the education of future experts in that domain. In addition, by examining the similarities and differ-

Correspondence should be addressed to Jarrod Moss, Department of Psychology, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15232. E-mail: jarrodm@cmu.edu

ences of expertise in one domain with expertise in other domains, a general understanding of the processes underlying expertise acquisition can be attained.

A number of domains of expertise have been studied, and there are some general findings about how the structure of domain knowledge changes with experience in these tasks. A hierarchical database of commonly occurring configurations or patterns is present in a number of domains, including games such as chess and Go as well as the interpretation of electronics diagrams (Chase & Simon, 1973; Egan & Schwartz, 1979; Reitman, 1976). This type of knowledge structure aids the expert in classifying this situation and identifying good moves. In particular, Egan and Schwartz proposed an explanation for their findings that they called *conceptual chunking*. This explanation says that the technicians were able to identify the conceptual category of a drawing by the layout of the wiring diagram (e.g., inverter, power supply, etc.). The conceptual category is organized in memory so that it is linked to a set of commonly occurring functional units, or chunks, that would normally appear in such a device, and then each functional unit is linked to a number of components that commonly instantiate that unit.

Another general property of expertise is that it affects not only what is stored in memory but also how things in the world are perceived and categorized. Experts have highly organized memory structures such as schemas, templates, and retrieval structures (Gobet, 1998; Richman, Staszewski, & Simon, 1995). As information about a new problem is perceived, this information automatically activates relevant domain knowledge and processes. This allows experts to easily categorize information and recognize solution schemas in their domain. For example, physics experts can recognize and solve common problems in a more efficient manner than novices through activation of appropriate strategies and schemas (Larkin, McDermott, Simon, & Simon, 1980). In fact, in expertise it is common for the associated knowledge structures and processes to be such an integral part of the cognitive system that they affect the way the expert perceives the environment. Chase and Simon (1973) argued that chunking can take place at the perceptual level, and elsewhere it has been shown that although domain knowledge does not directly transfer to other domains it can still influence the way people reason about general situations outside of their domain of expertise (Nisbett, Fong, Lehman, & Cheng, 1987).

Other general findings in the expertise literature are that experts tend to work forward from the givens in the problem rather than backward from the desired solution, and experts tend to classify items and problems in their domain of expertise according to a deeper conceptual structure rather than surface similarities (Chi, Feltovich, & Glaser, 1981; Hmelo-Silver & Pfeffer, 2004; Larkin et al., 1980; McKeithen, Reitman, Rueter, & Hirtle, 1981). For instance, experts in solving physics problems classify problems based on the principles used to solve the problem instead of similarities in the wording or diagrams of the problem (Chi et al., 1981). Computer programming experts are more likely to represent the functional similarities of keywords in a programming language rather than grouping words into phrases that are meaningful in English (McKeithen et al., 1981). These cases all involve possessing a representation of domain material that is adapted to key concepts that affect performance in a domain.

These properties of expertise are likely to hold in the complex domain of engineering design as well. Although it is worthwhile to examine the properties of expertise in this area, the studies reported here also examine the early acquisition of expertise, which has only been studied in a few domains. For instance, Hmelo (1998) provided an analysis of the acquisition of expertise in medicine. Her findings are consistent with what has been found in studies of medical expertise that compare professionals and students, but they also highlight some things that are unique to the early stages of expertise, such as the lack or incompleteness of advanced knowledge structures. For example, the medical students studied by Hmelo seemed to have a well-developed causal network of scientific concepts but lacked more advanced structures characteristic of experts, such as illness scripts. Understanding these knowledge structures and how they develop during the acquisition of expertise provides insights into the difficulties encountered by students due to their incomplete knowledge structures. This understanding may also highlight methods that can be used to encourage the development of more advanced and coherent knowledge structures. The studies presented in this article focus on the representation differences that exist among engineering students in the early stages of expertise acquisition.

Previous work examining differences in early expertise among engineering design students has focused mainly on how the design process changes with experience. For instance Atman, Chimka, Bursic, and Nachtmann (1999) looked at differences in the design processes by analyzing concurrent verbal protocols of freshmen and senior students solving a design problem. They found that seniors have a better representation of what stages a good design process includes and can transition between steps in the design process more easily than freshmen. Seniors also consider more design alternatives than do freshmen, which is probably one reason that seniors end up producing better quality designs in their study. Other researchers have examined differences between the design processes of students and professionals. For instance, it was found that in an artificial design task that groups of professionals exhibit more metacognitive and strategic behaviors during design (Smith & Leong, 1998). Student groups rarely exhibited these behaviors, and they tended to iteratively refine their original design concept as opposed to exploring multiple alternatives as the professionals often did.

Another series of studies investigated engineering design processes in students and professionals using concurrent verbal protocols (Ball, Evans, & Dennis, 1994; Ball, Evans, Dennis, & Ormerod, 1997; Ball & Ormerod, 1995). These studies found that novices use a depth-first design process, whereas experts use more of a breadth-first approach. This means that although both groups of designers decomposed the problem into modules, experts tended to develop each module to a certain level of detail before moving to the next level of detail, and novices were more likely to do detailed design on one module before moving on to the next. The authors proposed that this depth-first structure is advantageous for novices because it limits the number of unfulfilled goals they must store in memory. This is because it is not necessary to maintain detailed unfulfilled goals for partially complete modules. The depth-first design process they discuss is simpler because a module can be in only one of three states: completed, never been worked on, or currently being designed.

These studies have some things to say about the cognitive processes going on in design, but there is not much at all about the internal representations that engineers use while solving a design problem. Goel (1995) did present an analysis of design problem spaces and the results of a study that indicate that certain types of symbol systems are necessary to support design activity. In particular a system that supports design requires a level of ambiguity in its representation of an incomplete design. This work differs from ours in that it is concerned with some general properties of a representational system and not the specifics of how design information is organized in memory. Another difference is that the participants in his study were all experts, so it is unclear how these results map onto novices.

Understanding the representation changes that occur as engineering students progress toward becoming professionals is essential in achieving an understanding of the cognitive processes underlying performance in engineering design. It is well known that the representations that experts and novices employ while solving problems drastically alter the complexity of the problem space (Koedinger & Anderson, 1990; Newell & Simon, 1972). These changes are probably occurring during the early stages of expertise acquisition in a domain, which makes it important to study students at different points during the transition to expertise. The work presented here represents the beginning of a detailed examination of the representations and processes that allow engineering students to acquire the skills necessary to perform the complex tasks required by their profession.

One of the key concepts in engineering design is function. A study of a similarly technical field, electronics, found that the function of a device and subsystems within that device affected how it was represented by expert electronics technicians (Egan & Schwartz, 1979). Much of the engineering design research literature has assumed the importance of function in design problem solving (for a review, see Wood & Greer, 2001). This includes studies of human engineers as well as computational design systems. For instance, it has been shown that automated design systems benefit from representing the functions of sets of components (e.g., Moss, Cagan, & Kotovsky, 2004). These properties of the literature indicate that engineers perceive the importance of function within their domain. An additional motivation for the studies presented here is to investigate exactly what role function plays in the early stages of expertise. In particular it will be argued that the representation of function plays a key role in forming coherent representations of electromechanical devices.

Throughout the article a distinction is made between interactions of components and functions of components when examining what students write about devices. We differentiate between these two categories by assessing whether the statement in question includes only causal information about the way two components interact or whether more detailed information is included in the statement. There is a key difference in understanding between a student who states that “This gear turns that gear” and a student who might say, “The smaller gear turns the larger gear so that the larger gear turns slower than the smaller one” or one who might say, “The pair of gears serves to decrease rotational speed and increase torque.” We would call the latter two statements functional statements, whereas the first would be considered a statement about the interaction of two components. The functional statement need not contain technical terms as the second example statement shows, but it must contain some information about the function of components beyond a simple indication of causal interaction. This additional information usually indicated the purpose of a component or set of components in the system.

The studies presented here look at freshmen and senior engineering students to see what kind of representation changes accompany the early transition to expertise. The kinds of differences found between freshmen and seniors may generalize to professional engineers on further investigation, or alternatively the transition from student to professional may involve other qualitative changes in mental representation. Even if these differences do not generalize to professionals, the results still provide some insight into how expertise is acquired.

A type of hierarchical chunking structure was hypothesized based on a description of conceptual chunking by Egan and Schwartz (1979). They used the idea of conceptual chunking to explain how experts and novices represent devices depicted in electronics diagrams. This

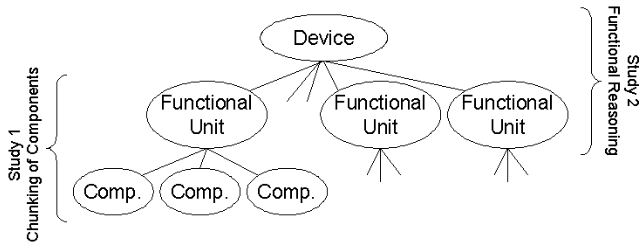

Fig. 1. One interpretation of the conceptual chunking explanation presented by Egan and Schwartz (1979).

structure is presented in Fig. 1 and essentially shows that components can be expected to be chunked into meaningful units based on the function they perform in the device, and these functional units then interact to produce the overall device behavior. Essentially, the middle level is a set of subsystems that the device can be broken down into, and each of these subsystems performs a particular function within the device. These subsystems can then be broken into finer units, such as a set of basic components where each component performs a simpler function. The two studies presented here focus on different levels of device representation as can be seen in Fig. 1. This first study examines some basic differences in how components in devices are represented and chunked together. The second study seeks to determine whether seniors think about and represent devices in a more abstract functional manner than do freshmen.

2. Study 1: Chunking of components

This study examines how the students chunk components into larger meaningful units. Just as experts are known to chunk elements into larger units of knowledge in other domains, it should be the case that the more experienced students have some way of organizing knowledge about components in a device. One hypothesis is that components will be chunked into larger meaningful units that perform a certain function in the device. Such functional units or chunks could occur across multiple devices in which the chunk performs the same function. One example would be a rack and pinion, as this set of components is one common method to convert between rotation and translation and could be expected to occur in a variety of devices.

To investigate these issues, a recall paradigm was used that extends an approach used by others to study chunking differences in expert–novice behavior (Chase & Simon, 1973; Reitman, 1976). The basic method is to present a stimulus, such as a chess board in a midgame position, for a brief period of time. The participant is then asked to recall the presented stimulus. In the original methodology both recall and perception tasks were used, and chunks were identified based on interresponse times (IRTs), which are pauses between actions, that were common to both tasks. However, later work examining chunks in Go (Reitman, 1976) found that a common IRT could not be found for both tasks due to the fact that the chunks in Go have an overlapping structure. This was not necessarily a problem in the chess research because the chunks in chess have more of a hierarchical relation. In our study, only a recall task was used, and to avoid problems with finding an appropriate IRT boundary, a number of different measures were used to examine representation differences. In particular, we looked at percentage recall after one exposure, errors, patterns of recall, and chunks. The stimuli being recalled were diagrams of electromechanical devices.

2.1. Method

2.1.1. Participants

There were 15 seniors majoring in mechanical engineering who volunteered for the study. These students were recruited from a senior engineering design course at Carnegie Mellon University. There were 15 freshmen engineering students who also participated in the study as partial fulfillment of a course requirement. All freshmen were enrolled in the engineering college at Carnegie Mellon, but students in this college do not declare a particular engineering major until after their freshman year.

2.1.2. Stimuli

Three electromechanical devices were represented in schematic diagrams that indicated how components connected together in each device. The three devices used were a drill, a pressure gauge, and a weighing machine (scale). The schematics were represented in an idealized fashion where only the type of each component and the connections between the components were displayed. Connections between components were represented by lines connecting components. An example design schematic is shown in Fig. 2. All of the devices contained 14 to 16 components and 14 to 15 connections. The number of unique components differed because some types of components were used more than once in a design. The drill had only 9 unique components, whereas the pressure gauge and weighing machine had 11 and 12 unique compo-

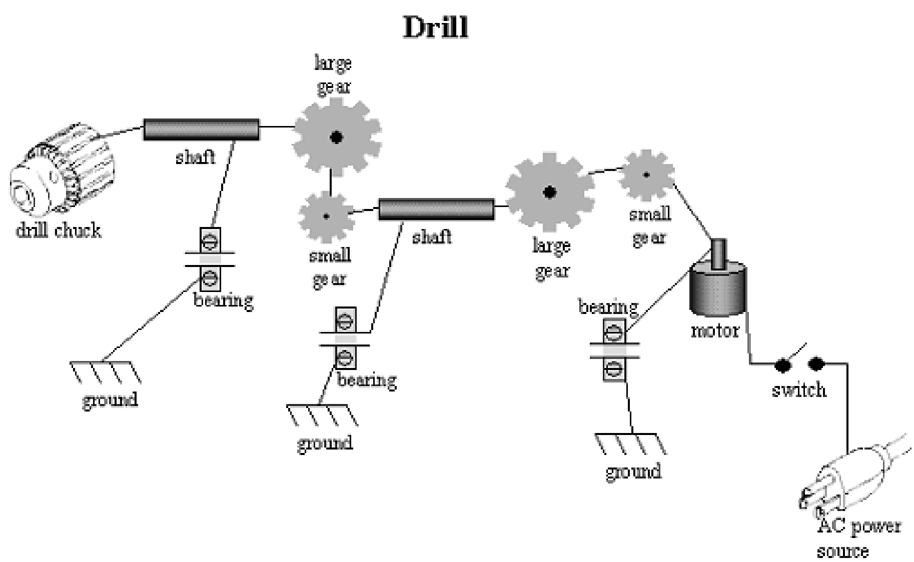

Fig. 2. Example of diagrams seen by participants.

nents, respectively. Each diagram also had a label at the top indicating the type of device as depicted in Fig. 2.

2.1.3. Procedure

Participants were asked to recall three design schematics, using a graphical interface after a brief study period. Participants first received instructions and were allowed to become familiar with the interface and the type of representation used in the diagrams. The user interface is depicted in Fig. 3, and it consists of a set of components that can be dragged over to a drawing space where they can be moved, connected, disconnected, or removed. Participants then received a practice trial followed by three recall trials. During each trial, the initial schematic was displayed for 40 sec, and then the display of the user interface replaced the schematic. Participants then had 3 min in which to recall as much of the schematic as possible. The design schematic was presented again for 40 sec if the participants had not recalled the design completely. These periods of display and recall alternated until the participant recalled the device perfectly. The task was done until perfection so that participants would be motivated to recall as much as possible in a minimum number of viewings given that they had to wait another 3 min to obtain another viewing of the device.

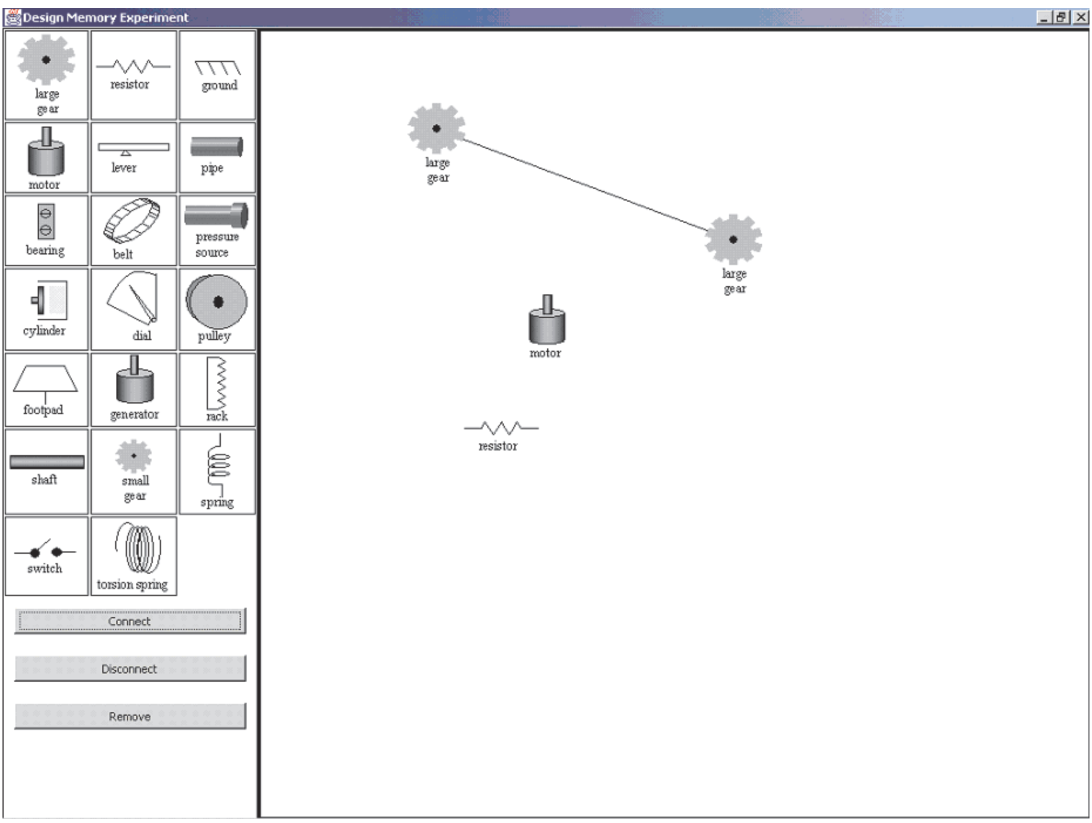

Fig. 3. Screen shot of user interface.

The presentation order of the three design schematics was counterbalanced. The computer generated a time-stamped entry in a log file for every action the participant took. The log was detailed enough so that a participant's actions could later be replayed. After the recall period participants were asked to indicate how they grouped components during recall by drawing circles around groups of components in the device diagrams. They were also asked to indicate why they grouped these sets of components together.

2.2. *Results*

The results are grouped into three categories. The first set contains quantitative results, including the percentage of the devices recalled after one presentation of the diagram and the number of errors made while recalling the devices. The next set of results details the patterns of behavior that were observed while recalling the device. The final set of results deals with how components of a device were chunked during recall. For all statistical results reported in this article, an alpha of .05 was used. When multiple comparisons were made, Bonferroni corrections were used.

2.2.1. *Percentage recall and errors*

The percentage of components and connections among components recalled correctly during the first recall session of a trial was analyzed using a 2 × 3 analysis of variance (ANOVA), with the three devices as a within-subjects factor. The average proportions of components and connections recalled after one presentation are presented in Table 1. Recall of components was almost perfect for both freshmen and seniors, and there was no main effect of experience level, $F(1, 28) = 1.1$, $p = .30$, or device type, $F(2, 56) = 1.8$, $p = .17$. Recall of connections was lower overall than for components. There was no significant main effect of experience, $F(1, 28) = 1.3$, $p = .26$, but there was a significant main effect of device type, $F(2, 56) = 5.63$, $p < .01$. Further comparisons revealed that the drill's connections were recalled better than those for the weighing machine, $F(1, 28) = 11.4$, $p < .01$, and the pressure gauge connections were not significantly different from either of the other two devices. There was no significant interaction between experience level and device for either the components recalled, $F(2, 56) = 0.7$, $p = .50$, or connections recalled, $F(2, 56) = 1.8$, $p = .17$. These results indicate that although freshmen and seniors performed equivalently, the weighing machine was the most difficult device to recall correctly, and the drill was the easi-

Table 1
Proportion of components and connections recalled after one presentation

	Drill		Pressure		Weigh	
	M	SD	M	SD	M	SD
Components						
Freshmen	.95	.07	.91	.12	.91	.09
Seniors	.96	.08	.95	.09	.94	.07
Connections						
Freshmen	.91	.12	.74	.24	.69	.27
Seniors	.88	.17	.88	.19	.79	.24

est. These differences are consistent with the idea that the difficulty of recall increases as the number of unique components in a device increases.

A number of error types were defined that include insertion, removal, connection, and disconnection errors. Insertion errors occur when a component is added to the device under construction and that type of component is not in the original device or when an added component is not needed because the device under construction already includes the correct number of that type of component. A removal error occurs when a component that is needed in the device has already been added but the participant removes it. A connection error occurs whenever the participant connects two components that are not connected in the original device. A disconnection error occurs when a participant has correctly connected two components in the device but then disconnects them. The error data are presented in Table 2. Freshmen made more errors overall than seniors, $F(1, 28) = 5.9$, $p = .02$, and there was also a main effect of device, $F(2, 56) = 4.1$, $p = .02$. Further comparisons show that more errors were made on the weighing machine device than on the drill, $F(1, 28) = 7.5$, $p = .01$. The interaction between experience level and device only approached significance, $F(2, 56) = 3.1$, $p = .054$.

The only type of error that occurred frequently enough to analyze separately was the connection error (see Table 2). There were effects of experience, $F(1, 28) = 5.2$, $p = .03$, and device on the number of connection errors, $F(2, 56) = 3.9$, $p = .03$. There was also an interaction of device and experience, $F(2, 56) = 5.0$, $p = .01$. Further comparisons show that freshmen made significantly more connection errors than seniors on the weighing machine device, $F(1, 28) = 9.6$, $p < .01$, but not on the other two devices. Freshmen made more connection errors on the more difficult device, whereas seniors consistently made less than one connection error per device. The connection errors account for the effects seen in the overall set of errors.

The high performance on the recall measure may indicate that the task was too easy for both groups of participants. This could be because the 40 sec for which the device was presented was too long and diminished the differences between the two groups. This possibility is discussed in more detail later. Even with this limitation, the difference in number of errors is reliable, and it is related to the chunking results discussed in the following section.

2.2.2. Chunking results

There are a number of methods that may be used to determine chunks from the data that were collected. One method would be to choose a criterion IRT that would be used to deter-

Table 2
Average number of errors and connection errors

	Drill		Pressure		Weigh	
	M	SD	M	SD	M	SD
Total errors						
Freshmen	1.53	2.17	2.13	3.14	4.87	4.82
Seniors	1.20	1.61	1.13	1.55	1.40	1.92
Connection errors						
Freshmen	0.87	1.12	1.33	1.88	2.80	2.51
Seniors	0.80	1.32	0.73	0.88	0.67	0.90

mine chunk boundaries (e.g., Chase & Simon, 1973). In addition to the IRT technique, other methods such as constructing lattices of chunks (Reitman & Rueter, 1980) and a type of hierarchical clustering (Baggett & Ehrenfeucht, 1988) have been used to determine conceptual organizations in memory. These techniques are not applicable because the lattice construction technique requires a participant to recall one device multiple times, and the clustering technique assumes that chunks only consist of connected components. However, it was possible to use another version of hierarchical clustering on the recall times of components. One advantage of this technique is that it allows for chunk boundaries to be defined without having to identify a criterion IRT to distinguish chunk boundaries. The only requirement is that recall within a chunk should generally be faster than recall between chunks, which has been found to be true in other work on chunking (Reitman & Rueter, 1980).

The clustering technique for defining chunk boundaries starts with the times at which each component was added during the first recall attempt. A traditional single-linkage hierarchical clustering algorithm was used to define chunk boundaries. This type of clustering algorithm works by looking through the set of items and placing the two that are closest, or most similar, into a single cluster. This cluster is then added to the set of items as one item, and the distance between it and any other item is defined as the distance between the other item and the closest item in the cluster. This is repeated until all of the items have been consolidated into one large cluster. The end result is a hierarchical cluster of items that is often represented in a tree structure as shown in Fig. 4. In this study, the *distance between any two components* is defined as the difference in time between when the two components were recalled. This algorithm creates a

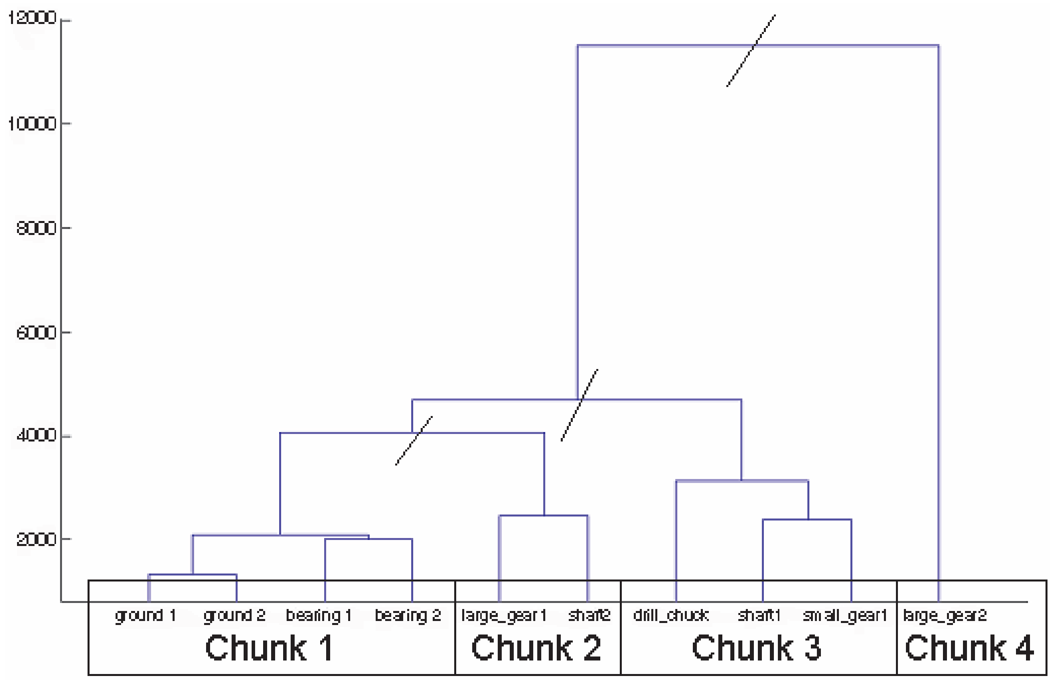

Fig. 4. Example of clustering method.

hierarchical tree structure in which the leaves are the components, and one example of the hierarchical tree for the drill device is shown in Fig. 4. Each branch, or link, in the tree connects two clusters, and the height of the link in the figure represents the distance between the two components. Chunks are identified by determining which links in the tree should be cut to define the chunks.

To determine which links to cut, an inconsistency score for each link is calculated, based on the distance between the clusters of components that the link connects. The inconsistency score incorporates information about all of the links below this one in the tree, including the average and standard deviation of the distances for each of the links below this one in the tree. It essentially calculates a z score for this link, based on all of the links below this one in the tree. The average of all of the distances of the links below this one is subtracted from this link's distance, and the result is divided by the standard deviation of all of the distances of the links below this one. Because our metric of distance between two components is the elapsed time between when they were each recalled, this inconsistency score is essentially a way to make a judgment about whether the time difference is significantly larger than other time differences for a particular participant. Large time differences are likely to be between chunk transitions, whereas smaller ones are more likely to be within chunk transitions.

In these analyses, an inconsistency score above 0.9 meant that the link was cut to form two separate clusters. Inconsistency values above 1.0 yielded a few large chunks, whereas values smaller than 0.8 yielded many small chunks. A value of 0.9 was used because it was in the center of the range from 0.8 to 1.0, and it produced between four and nine chunks for each participant. This seemed like an appropriate number of chunks, based on past research on chunking, which indicates that the number of chunks is consistent with known working memory constraints for short-term recall of common words (e.g., Chase & Simon, 1973). An example of the result of this process is illustrated in Fig. 4 where four chunks are defined by cutting the appropriate links. As long as there is some consistent difference between within-chunk and between-chunk pauses, this technique will have the desired effect. The chunks identified by the clustering in Fig. 4 are shown in the drill diagram in Fig. 5. As can be seen, this person only recalled 10 components of the device in the first trial.

After the chunks had been determined using this clustering method, the chunk sizes for freshmen and seniors were examined, and they did not differ. Chunks contained 1 to 5 components with the overall average chunk size being 2.15 components per chunk. Qualitative differences between the chunks identified for freshmen and seniors were identified by examining the frequency with which specific chunks were used by both freshmen and seniors. Only chunks that were used by at least 5 of the students were included in the analysis so that the chunks examined were chunks that were commonly agreed on and were not unique to just 1 or 2 students. For each chunk, the proportions of freshmen and seniors who used that chunk at least once were calculated.

There were many similarities between the chunks used by both freshmen and seniors, but there were some notable differences in the proportion of students who used a particular chunk. The proportions of students who used a particular chunk are presented in Table 3. The chunks used more by seniors than freshmen appear at the top of the table, and the chunks used more by freshmen appear at the bottom. In general, the chunks used more by seniors tend to have a meaningful function within the device. For instance, the large gear–rack chunk converts rota-

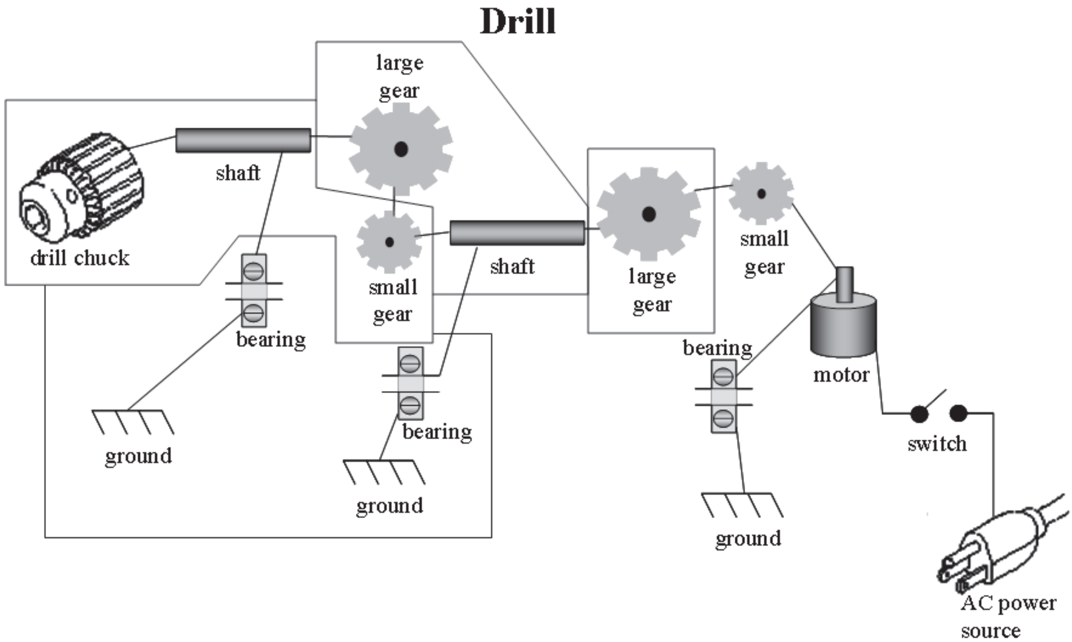

Fig 5. An example of chunks produced by clustering technique.

tional motion to translational motion, the footpad–cylinder–cylinder and the cylinder–cylinder chunks reduces the input force, and the motor–resistor chunk provides dampening. The chunks used more by freshmen included chunks of the same type of component and other chunks with less functional meaning. The one notable exception to this general pattern was the large gear–small gear chunk, which serves to decrease velocity and increase torque. Overall, this analysis indicates that freshmen and seniors shared some of the same chunks, but the seniors were more likely to use chunks with a functional meaning.

As an additional way to assess whether the chunks identified by both groups were functional in nature, an expert was asked to identify groups of components that performed a particular function in the devices that were used. The expert identified 4 to 5 chunks for each device. For example, the expert identified 5 chunks for the drill device, which included each pair of gears, each shaft-bearing-ground set, and the motor–shaft–ground chunk. The expert stated that the gear pairs increased torque, and the other chunks all served to restrict motion in some directions. The number of times that the expert chunks were used by the students was identified and analyzed using a 2×3 ANOVA with device as a within-subjects factor. There was a significant difference between freshmen and seniors, $F(1, 28) = 6.37$, $p = .02$, with seniors matching 1.2 of the expert chunks on average and freshmen matching .78 chunks. There was also a significant effect of device, $F(2, 56) = 6.49$, $p = .003$, with there being fewer matches on the weighing machine than the other devices.

The groups of components that the students circled after the recall task and their reasons for grouping those components were also analyzed. The reasons written by students were exam-

Table 3
The most frequent chunks identified by freshmen and seniors

Chunk	Freshmen	Seniors
Large gear–rack	.13	.40
Footpad–cylinder–cylinder	.33	.53
Motor–resistor	.07	.27
Cylinder–cylinder	.27	.40
Large gear–small gear–shaft	.20	.33
Shaft–bearing–ground	.20	.33
Lever–lever	.20	.33
Power source–switch–motor	.13	.20
Bearing–ground	.40	.40
Pressure source–cylinder	.27	.27
Spring–ground	.20	.20
Ground–ground	.20	.20
Ground–ground–ground	.40	.33
Drill chuck–shaft	.20	.13
Large gear–small gear	.67	.60
Bearing–bearing–bearing	.40	.27
Dial–torsion spring	.40	.27
Cylinder–lever–lever	.40	.27

ined for any mention of the function of a group of components. As discussed earlier, there was a distinction made between talking about the function of a group of components and the interactions of those components. The participants were sorted into groups based on whether they gave a functional reason for a grouping at least once in any of the devices. No freshman gave a functional reason for any grouping, but 11 of the 15 seniors did, $\chi^2(1, N = 30) = 17.37, p < .001$. A second person was also asked to code the responses, and there was moderate agreement, Cohen's $\kappa = .67$. Essentially the second coder had a lower criterion for coding a chunk description as functional and included 3 additional seniors as well as 2 of the freshmen in the function category. This did not affect the pattern or significance of the results. The freshmen responses included some descriptions of interactions, but they mainly gave reasons dealing with visual salience or ease of grouping names of components into some mnemonic. These results further support the idea that the seniors were thinking about the functioning of chunks of components more than were the freshmen.

The chunks were also analyzed in relation to the connection errors by looking at where each participant's connection errors occurred relative to that individual's chunks. For both groups, 84% of the connection errors occurred when connecting two components that were not in the same chunk. In addition, the proportion of connections falling between chunks did not differ for freshmen and seniors, $F(1, 28) = 1.68, p = .21$, with freshmen and seniors having 64% and 61% of their connections fall between chunks, respectively. However, the absolute number of these between-chunk connection errors was quite different for the two groups. Freshmen made 62 of these between chunk connection errors, whereas seniors made only 29. This means that the hardest connections to remember for both groups were the connections between chunks, but freshmen had significantly more difficulty remembering these between-chunk connections.

2.2.3. Patterns of recall

The order in which components were recalled was also analyzed. After one component is recalled, the next component recalled can either be connected to the previous component or not. Recall that is highly connected demonstrates the importance of connections in the representation and recall strategy used by the participant. Connectedness was examined by looking at the number of breaks in recall, where a *break* is defined as consecutively placing two components that are not connected. In all of the designs to be recalled, a certain number of breaks were required to complete the design due to parallel connections, but unnecessary breaks indicated that the student was not following the interconnected structure of the device during recall. For instance in the drill shown in Fig. 2, three breaks are required to handle the three bearing–ground elements. The number of breaks above the minimum required was used in this analysis. Freshmen made significantly more breaks in recall than did seniors, $F(1, 56) = 4.84$, $p = .03$, as can be seen in Table 4.

The high degree of connectedness for seniors' recall could mean that recall was done by recalling successive components that shared a connection. However, an alternative explanation is that spatially proximal components have a good chance of being connected, so spatial proximity could be the main factor in determining the pattern of recall. To examine this possibility, the proportion of times participants transitioned from one particular component to the next was examined. Fig. 6 shows a representative sample of the results of this type of analysis on a portion of the drill for seniors. Only proportions greater than or equal to .2 are displayed in the figure as these correspond to the transitions made by at least two participants. With few exceptions the majority of transitions occurred along connections, rather than to unconnected but spatially proximal components. Almost all of the exceptions occurred in transitions dealing with groups of the same type of component (bearings and grounds in Fig. 6). The fact that a number of spatially proximal but unconnected components were not recalled consecutively is evidence against the use of spatial proximity in the representations and recall strategies. For example, the small gear on the left side of the drill in Fig. 6 is near five different components, but 93% of the time participants transitioned to only the connected components (the shaft and large gear). Some of the exceptions can be explained by noting that a number of participants added all three bearings from left to right followed by the three grounds from left to right. This is apparently just a convenient pattern to recall this set of similar components, and it does not seem to offer evidence for a spatial proximity explanation.

The expected proportion of transitions from one component to another was calculated by dividing the number of components a component was connected to by the number of components that it was adjacent to in the diagrams. This expected proportion was compared to the ac-

Table 4
Average breaks in reconstruction (above minimum)

	Drill		Pressure Gauge		Weigh	
	M	SD	M	SD	M	SD
Freshmen	3.73 (2.60)	2.60	2.80 (2.14)	2.14	5.67 (3.15)	3.15
Seniors	2.73 (2.19)	2.19	1.60 (1.24)	1.24	3.47 (1.96)	1.96

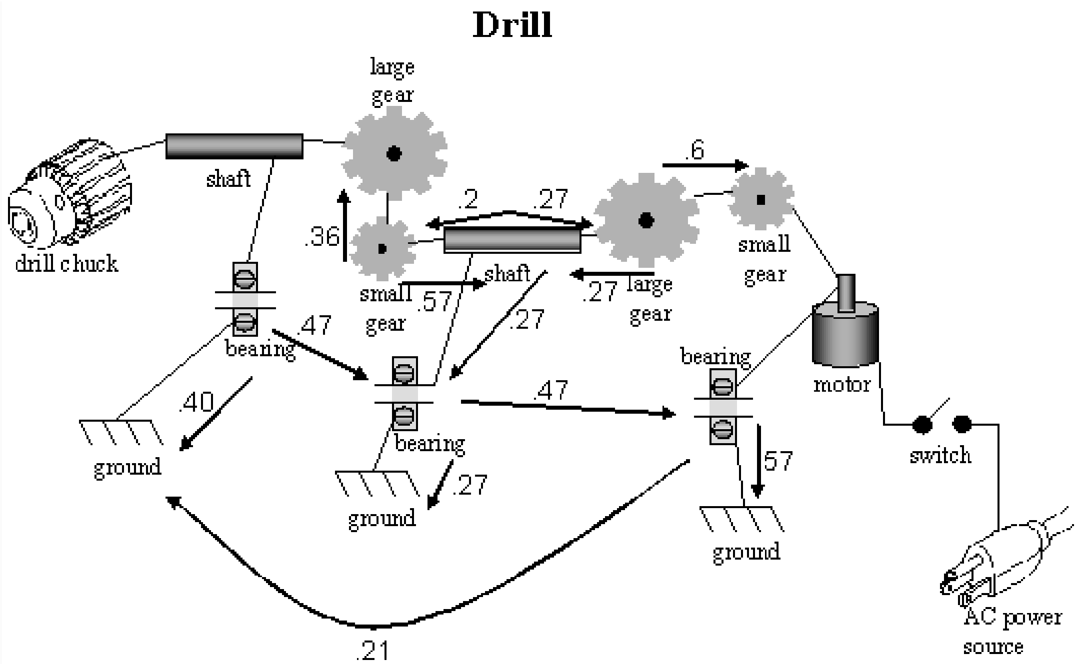

Fig. 6. Example of the percentage of transitions from one component to the next.

tual proportion of transitions to connected components for all of the components in all of the devices. The components that were only connected to one other component were excluded because these components were often at the edges and, once they were transitioned to the only transition away, involved an unconnected component. The expected and actual proportions were compared using a Wilcoxon signed ranks test. There was a significant difference in these proportions for seniors, $z = -2.53$, $p = .01$, with seniors making 72% of their transitions between connected components as compared to the 62% expected. Freshmen made 64% of their transitions between connected components, which was not greater than expected, $z = -.887$, $p = .38$. The high degree of connectedness for seniors is therefore not due to spatial proximity alone.

Another pattern that was evident in the order in which components were recalled was that a number of students started at the input of the device and reconstructed the device, based on the flow of energy through the device. For example, in the drill in Fig. 2, students following this pattern began with the power source and then proceeded to add the switch, motor, gear sets, and drill chuck in that order. This pattern was formalized by defining it as starting with the input and making three or fewer breaks in connectedness during recall. This pattern was used at least once by 14 out of the 15 seniors and by 8 of the 15 freshmen. A chi-square test shows that this difference was significant $\chi^2(1, N = 30) = 3.97$, $p < .05$. This provides some evidence that many seniors preferred to reconstruct the device, based on the flow of energy through the components of the device. Freshmen were less likely to reconstruct the device in any such recognizably meaningful order.

In summary, the recall patterns observed emphasize the importance of connectivity in the representation of the devices. Seniors are more likely to follow the energy flow through a device's connections from input to output, and they are more likely to recall successive components that are connected. This connectedness cannot be explained by spatial proximity alone. These patterns indicate that the students at least understand the flow of causal behaviors within the device, and this understanding is of course a necessary component to understanding the function of a component or a group of components within the device. These findings when combined with the chunking results and the difference in connection errors provide a basis for understanding the differences in the device representations used by the two groups.

2.3. Discussion

The results indicate that the students do decompose devices into smaller chunks, and one of the main reasons that components are chunked together is because of the function they perform as a group. Freshmen are less able to recognize functional groups of components, and they are less able to fit together the chunks they do use. This is shown by the fact that freshmen make more connection errors, and these errors occur predominantly at chunk boundaries. The fact that the chunks that seniors identify are more likely to be functional units can be seen from the analysis of their chunks, including comparison to an expert and the reasons they wrote for grouping the chunks.

Another result that emphasizes the importance of functionality and connections in device representations is that seniors recall devices in a more connected manner as measured by the number of breaks in recall. They are also more likely to follow the energy flow through the device during recall. This could be indicative of a structured traversal of a coherent mental representation of the given device. As discussed in the introduction, previous work indicates that a hierarchical chunking structure such as the one in Fig. 7 is a likely structure for storing device knowledge. If the representation of the device is a hierarchical structure such as this one, then the strategy of following the energy path through the device during recall is accomplished by a specific traversal of the device representation. For instance, if one starts with the device and follows a link to the first functional unit of the device, then the first step is to recall the components that make up that unit while remembering how they are connected. The next step is to go back to the functional unit node and follow the link to the next functional unit. Even if the link to the next functional unit is not recalled correctly, it may still be possible to reason about what

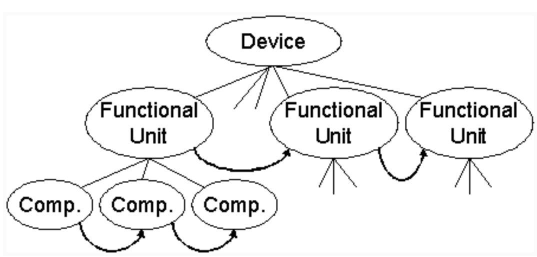

Fig. 7. One possible hierarchical chunking representation.

the next functional unit may be, based on the person's understanding of the overall device behavior and the functional units already recalled. This type of representation and recall process requires that the connectivity between and within chunks be adequately represented.

One possibility is that explicit links are constructed in the representation such as those between functional units in Fig. 7. Another possibility is that the participant follows a specific encoding process such as encoding the input chunk as the first functional unit and then the next set of components in the device's flow of energy as the second functional unit and so on. However relying strictly on the ordering of functional units in the device would make it difficult to deal with parallel connections, and the participants did not demonstrate any particular difficulty with parallel connections.

The results from this study imply that the freshmen are not necessarily decomposing the device into functional units, and instead they may be creating chunks based on visual similarity, proximity, or through a mnemonic strategy. This is apparent in the fact that their chunks relate less to the function of the device than do those of seniors. This finding is not unexpected because it has been shown that novices often use concrete surface features, whereas more abstract deeper features are preferred by experts (e.g., Chi et al., 1981; Feltovich, Spiro, & Coulson, 1989). It also explains why they have difficulty in remembering connections between chunks as they do not have the functional relations between chunks to rely on as the seniors do. The finding that their recall is less connected and less likely to follow the flow of power through the device would also be expected if their representation does not incorporate functional chunks and interconnections.

One limitation of this study is that devices may have been presented to participants for too long, so participants found the recall tasks fairly easy as shown by the high recall performance after just one presentation of the device and the low number of errors. Participants may have used other strategies than just chunking because the device was presented for such a relatively long period of time (40 sec). However, in a study of chessboard positions, which used the same recall paradigm while manipulating presentation time from 1 to 60 sec, it was found that participants remembered more chunks with longer presentation times, but there were no qualitative differences in cognitive processes for the longer presentation times (Gobet & Simon, 2000). They found that chess experts would occasionally combine smaller chunks into larger chunks or remember a larger number of chunks, but individual pieces were not simply added to existing chunks. This leads us to believe that the presentation time is not a serious issue and that, with the exception of chunk size, the differences between freshmen and seniors would have been qualitatively similar with a shorter presentation time. The long presentation time may be the reason that we did not find a difference in chunk size. Freshmen would have had time to form chunks, using more time consuming mnemonic strategies, whereas seniors would have had time to encode the few small single-component chunks that did not fit with their larger chunks. This would have decreased the average chunk size for seniors but increased it for freshmen.

The results of this study illustrate some key differences between these sets of students in the early stages of acquiring expertise. Both groups are able to chunk components into larger units, but the seniors rely more on functional information to create a coherent mental representation of a device. The next study was designed to provide additional information on the importance of the function of groups of components in the mental representations of engineering students.

3. Study 2: Functional reasoning

The purpose of this study is to determine if a kind of abstract functional understanding was more prevalent in the mental representations of seniors. Because it is apparent from the first study that freshmen lack a strong representation of how chunks of components connect and function, this should mean that freshmen are not as able to reason about devices in an abstract functional manner because this ability would depend on being able to reason about the functions of chunks of components and how those functions connect and interact in producing device behavior. Another purpose of this study is to see if experience in the domain will lead seniors to produce representations that are more similar to one another than those freshmen do.

Other work has demonstrated that people high in self-rated mechanical ability appear to reason better about the functioning of a device than do people low in self-rated mechanical ability (Heiser & Tversky, 2002). In this study it is assumed that the participants are all high in mechanical ability because they are all majoring in or intending to major in mechanical engineering. Based on these results and those of Study 1 it was hypothesized that more experienced participants would focus more on functional knowledge of a device than less experienced participants. To investigate these issues, a new method of data analysis is introduced. In the work by Heiser and Tversky, each proposition that was written by a participant was coded as either structural or functional. This allowed them to show that high-mechanical-ability participants used more functional propositions. In this study, latent semantic analysis (LSA) is used to test for higher functional content in written text. This method does not require someone to subjectively decide whether each proposition contains functional information or not. The data were also analyzed using a proposition coding system as well to show that the two methods are consistent.

Students were asked to write brief descriptions of devices that were presented in diagrams. One assumption is that the information students choose to include in a brief description is what they find important about the device and that this importance is related to their mental representation of the device. In addition to the issue of functional information discussed previously, another hypothesis that is investigated is that seniors are more mutually consistent in their descriptions than are freshmen. The reasoning behind this idea is that seniors have gone through years of formal education, which may lead them to think about the devices in a similar manner.

3.1. Latent Semantic Analysis

The participants' descriptions were analyzed using LSA. LSA was originally developed as an information retrieval technique designed to overcome synonymy problems (Deerwester, Dumais, Furnas, & Landauer, 1990). It has also been used for a number of other purposes, including as a model of text comprehension (Landauer & Dumais, 1997). More recently it has also been used to develop similarity metrics to be used in the analysis of data from complex problem-solving trials (Quesada, Kintsch, & Gomez, 2002). LSA begins with a word-by-document frequency matrix where the number of times each word appears in a document is recorded. This matrix can be interpreted as a multidimensional space where each column is a document vector, and each word is a row vector. For example, a document would be a column of numbers indicating the number of times each word appeared in the document. Each of the numeric entries in the document column can be seen as the coordinate of the document vector in one dimension of the multidimensional space. LSA reduces the dimensionality of this space to yield a multidimensional semantic space where each document or word is represented as a vector in that space. Through this dimensionality reduction technique LSA can provide a good measure of semantic similarity based on the shared context of two words. For example, the words *physician* and *doctor* may not appear in the same document, but LSA may conclude that they are similar, based on the fact that both words occur with words such as *patient, hospital,* and *nurse.* Similarities in this multidimensional semantic space can be computed between any two words or documents by computing the cosine of the angle between the vectors representing those two words or documents. The cosine is a measure of how closely aligned in meaning any two words or documents are. It ranges from –1 to 1, with smaller angles receiving values closer to 1, which indicates a high degree of alignment and similarity. However, negative values do not necessarily indicate that items are dissimilar in terms of semantic meaning, but negative and near-zero values do indicate that the items do not occur in the same contexts. For example, the word *hot* may occur in many of the same contexts as the word *cold,* and the two words would receive moderate to high cosine values, but the words *planet* and *tissue* would receive negative or near-zero cosine values.

One set of researchers used LSA to try to identify shared-design understanding among a set of designers (Hill, Song, Dong, & Agogino, 2001). That study was concerned with building information management tools that retrieved relevant information, based on a shared understanding of the design problem. This shared understanding was determined by using LSA to analyze documentation from design projects. Their use of LSA allowed for a representation of the desired design project to be created, but unlike this study, they were not concerned with identifying properties of this representation.

Dimensionality reduction techniques such as LSA are able to capture a number of the key similarities and differences between documents in a given set. These properties make LSA an excellent tool for exploring similarities and differences between documents written by participants, thus highlighting differences in the content of their representations.

3.2. Method

3.2.1. Participants

The participants were 44 volunteers from a senior mechanical engineering design class and 24 freshmen volunteers from a freshman mechanical engineering class. The freshmen participants were taking their first mechanical engineering course when the study was run.

3.2.2. Stimuli

Three electromechanical device diagrams were used in this study. These diagrams were taken from patents for a power screwdriver (Fig. 8), a cordless weed trimmer, and a drum brake system. The diagrams were mostly cross-sections of the devices and had lines labeling key components. The diagrams were used exactly as they appeared in the patent except that some excess and duplicate labels were removed. Each diagram had 9 to 11 labeled components and had the name of the device printed in large bold letters at the top.

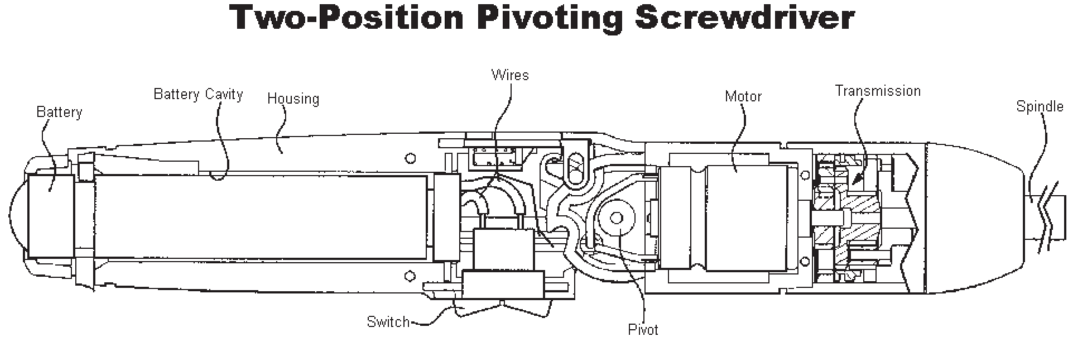

Fig. 8. Example diagram seen by participants (figure adapted from Alsruhe, 2001).

3.2.3. Procedure

Participants were told that they would see diagrams of three electromechanical devices that had been taken from patents. They were told that their task was to write a description of each device, but they were not told what kind of information to include in their descriptions. If they asked what to include, they were told to include whatever they thought important as long as it pertained to the device shown. Participants viewed the device on a computer screen and were told that they could click a button beneath the diagram that would remove the diagram and take them to a text area where they could type their description. There was also a button below the text area to take them back to the diagram, and they could alternate back and forth between description and diagram as often as they wanted. Each time they switched between the two views, an entry was added to a log file, and this state of their description was saved to a time-stamped file. Participants were instructed to spend about 5 min describing each device. They were not forced to spend exactly 5 min on each device but had to pace themselves to finish all three descriptions in 18 min. There was a clock displayed in the lower right corner of the screen to help them pace themselves.

After the descriptions had been written, the participants were asked to rate their prior knowledge of each device (1 = *poor,* 7 = *good*). The freshmen participants completed an additional set of eight true–false questions for each device before they gave their ratings. This set consisted of a mix of questions that emphasized either the structure–composition of the device or the function of components within the device. They were not allowed to view any of the diagrams while answering these questions. The purpose of these questions was to assess whether freshmen were able to access both functional and structural information from the diagrams. Seniors have used these types of cross-section displays before so it was assumed that they would be able to interpret them. The ratings were done after the questions because the ratings would involve presenting all three devices again, and the questions were intended to assess whether the participants could recall the structural–functional information without an additional presentation of the devices. The presentation order used here does mean that answering the questions could have influenced the prior knowledge ratings for the freshmen.

3.3. Results

Four seniors were excluded from all analyses because they failed to finish in the allotted time. One freshman and 5 seniors were also excluded because they either spent less than 3 min or greater than 7 min on a particular description.

The ratings for different devices for each group of students were analyzed using Friedman's test for related samples. Freshmen rated themselves as having more knowledge about some devices than others, $\chi^2(2, N = 21) = 7.3$, $p = .03$, with the weed trimmer being rated higher than the screwdriver, $\chi^2(1, N = 21) = 5.3$, $p = .02$. There were no significant differences between devices in the seniors' ratings. The rating differences between groups of subjects for each device were analyzed using the Mann–Whitney U test. There were no significant differences between the ratings of freshmen and seniors for any of the devices. The true–false questions were included to assess whether freshmen processed and remembered functional and structural knowledge about a device. All freshmen performed well on these questions, averaging 6.8, 7.2, and 7.2 questions out of 8 correct on the brake system, screwdriver, and weed trimmer, respectively. Any differences observed in functional knowledge were therefore not due to the freshmen being unable to access this knowledge.

One parameter that can be adjusted in LSA is the number of dimensions retained in the multidimensional space. Based on judgments from the number of dimensions used in previously published work with LSA, it was estimated that a good number of dimensions would be in the range of 50 to 300. LSA work that has been done with much larger text corpora has found that optimal dimensionality was around 300 dimensions (e.g., Landauer & Dumais, 1997). Because our corpus of device descriptions was much smaller, fewer dimensions were needed to capture most of the important information in the semantic space. The first 100 dimensions were used for all of the LSA results reported here.

The hypothesis that seniors are more consistent as a group than are freshmen was tested by computing a similarity measure between each participant's description of a device and the average vector for that device. The average vector was found by averaging the individual vectors for documents describing a particular device. One way to think about this average is as a type of prototypical description. A separate average was produced for freshmen and seniors for each device. For instance, all freshmen brake-device vectors were averaged to produce the average freshman brake description. Then for each device the average freshman vector was subtracted from the average senior vector. This produces a vector for each device that points from the average freshman description to the average senior description. Each participant's description for a particular device was then orthogonally projected to a location on this vector. The description's projected location along this vector provided a way of examining freshmen and senior differences. This method of comparing freshmen and seniors is more complex than just comparing distances in the original multidimensional space, but it provides a way of identifying the subset of dimensions most relevant for the purpose of distinguishing freshmen and seniors.

Due to the way the vector is constructed and the way documents are projected to points on this vector, the average senior description for a device will be a certain distance away from the average freshman description. The first test was whether there was a significant difference between where the senior and freshmen descriptions fall on this line, and they were significantly different, $F(1, 54) = 118.1, p < .001$. The hypothesis that seniors are more similar to one another as a group than freshmen was then tested by looking at how far freshmen and seniors are from the average freshman and senior descriptions, respectively. For instance, seniors should be more tightly clustered around the average senior description than freshmen are around the average freshman description. This difference was also significant, with seniors deviating less from their average description than freshmen from theirs, $F(1, 54) = 23.7, p < .001$. This supports the idea that seniors are more likely to be consistent with one another than are freshmen on the information they include in descriptions of a particular device. This finding is similar to what has been found in other domains where experts are more alike than are novices (McKeithen et al., 1981).

The search engine qualities of LSA were used to examine the hypothesis that seniors included more information about the function of a device in their descriptions. The documents in the multidimensional semantic space can be compared to a query vector, and their similarity to this vector can be assessed using the cosine measure. To formulate a query that represents function information, a set of words that are associated with describing the functioning of a device were combined into a single query. Stone and Wood (2000) developed a vocabulary to explain the chain of functions that produces a device's behavior. They showed that this vocabulary can be used to represent a variety of different devices. The function words they use and the associated list of synonyms that they define for those words totals 73 words (e.g., *actuate, adjust, control*). Three of these words were judged to deal more with the structure of devices, and they were excluded from the query. These words were *connect, locate,* and *join.* The remaining 70 words were combined into a query that was submitted to the LSA space.

A cosine score was generated for each document according to its similarity to this query of function words. The average cosine values are displayed in Table 5. Both experience level, $F(1, 54) = 5.3, p = .03$, and device type, $F(2, 108) = 8.9, p < .001$, had significant effects on a document's similarity to this query, and there was no interaction between these two factors. Further contrasts revealed that the weed trimmer descriptions had higher cosines than both the drum brake, $F(1, 54) = 13.7, p < .001$, and the screwdriver, $F(1, 54) = 13.0, p = .001$. The fact that descriptions written by seniors were more similar to the function words than those written by freshmen indicates that seniors included more content that is similar in meaning to the function words in the query. The magnitudes of the cosines are not very high because for any particular device description there are many function words that it does not contain. In this case the relative values of the cosines are more important than their magnitudes. It is the relative similarity to the function words that we are examining. Two examples of the descriptions that were used in the LSA analysis are presented in Fig. 9. One is a description that was ranked as highly simi-

Table 5
Average similarity to function words vector, using the cosine measure

	Brake		Screwdriver		Trimmer	
	M	SD	M	SD	M	SD
Freshmen	.13	.06	.13	.07	.18	.07
Seniors	.17	.09	.16	.07	.21	.09

High-function (senior)

The battery slides into the battery cavity where it comes into contact with the battery terminals. The terminals are connected to wires that run to the switch and then to the motor. When the switch is in the ON position the circuit will be complete between the battery and the motor. The pivot allows the user to change the shape of the screwdriver. The motor is on the upper half of the screwdriver so as to allow for this motion. The transmission is most likely a planetary gear system so that the screwdriver can run at various speeds and torques. The spindle connects the bit to the transmission/motor setup. Everything is contained by a housing in order to shield the user from the parts and to give the product a finished look.

Low-function (freshman)

This screwdriver is run off a very high RPM motor that transfers that goes through a transmission and creates larger a much larger torque [*sic*], but severely slows down the speed . It was designed so that the components of the handle are primarily the power source and switch, the motor on the other hand was put past the flexible part in order to do away with having to make some sort of flex joint that would be more expensive and would also take away some of the power. From a person stand point, these electronic screwdrivers are a piece of junk, they simply don't have the output that is needed for 95% of the work you need them for, nothing comes close to a good 18V electric drill.

Fig. 9. Examples of device descriptions.

lar to the function words query, and the other one is an example of a description from near the bottom of the rankings. It should be noted that even the low function description still includes some function information; however, it differs by including more information about other aspects, such as the location of components, and less function information than the high-function description.

Because this method of assessing function content has not been used previously, we also compared it to a method of assessing function content that involves coding propositions for function content used by Heiser and Tversky (2002). They essentially coded each proposition in a description as either expressing function or structure information. We elaborated on their coding scheme by breaking their structure category into categories we labeled *location* and *properties.* Location information describes how components are connected or where they are in a device. Properties of a device or components (e.g., size, shape, length) were included under the properties category. Their function category was broken down into two categories that we labeled *interaction* and *function,* which is the same distinction used in Study 1. Again, interactions were generic descriptions of component interactions that only conveyed which components were acting and possibly some causal information. Functions were descriptions of component interactions that were more detailed than actions, or they were statements that indicated the purpose of a component or set of components in the system. For instance, functions could differ from actions by indicating duration of motion, purpose of the component within the device, or other qualities beyond the basic fact that the components interacted in some way. In general, actions indicate knowledge of the causal behaviors in a device, whereas functions indicate knowledge in addition to causal behavior. There was also an “Other” category for statements that did not fit in the other four categories.

Table 6
Average proportion of propositions in each coding category

	Property		Location		Action		Function		Other	
	M	SD	M	SD	M	SD	M	SD	M	SD
Freshmen	.18	.08	.22	.13	.22	.10	.24	.09	.15	.16
Seniors	.17	.11	.21	.14	.20	.11	.32	.15	.11	.19

The descriptions were broken into propositions and coded into one of the four categories in the coding scheme. A random subset of the descriptions (one third of the total) was coded by an independent coder, and there was 86% agreement between the two coders, Cohen's $\kappa = .82$. The proportion of propositions coded into each category is given in Table 6. It can be seen that the largest difference between freshmen and seniors was in the function category. The proportion of propositions that were classified as function should be a measure similar to that computed by LSA. Seniors had a significantly higher proportion of function statements in their descriptions than did freshmen, $F(1, 54) = 5.79$, $p = .02$. The proportion of function statements is also correlated with the LSA cosine score, $r(56) = .71$, $p < .001$. These results further support the finding that seniors think about function more than freshmen, and they also indicate that the more automated LSA analysis agrees fairly well with a more traditional coding-scheme analysis.

3.4. Discussion

The results of this study support the idea that function information is more prominent in the representations used by seniors, and there is also support for the idea that seniors employ more similar representations than do freshmen. The finding that freshmen could answer questions about the function of components in a device means that they were capable of understanding the function of a set of components when asked to, but they did not include as much function content as seniors in the device descriptions they generated. This is a further indication that function was not as prominent in their representations as it was in those of seniors. It could indicate that for seniors function was the glue that bound the different parts of the device together, whereas for freshmen it was an additional part of a more fragmented representation.

These findings are consistent with the general notion that the main difference between the two groups of students is the ability to represent the functionality of chunks of components in a device and how these functional units interact to produce device behavior. This representational advantage is one reason seniors were able to minimize errors in the device recall task. Focusing on the function information may have allowed seniors to form a coherent representation of the device, and this common focus is likely the reason that seniors were more similar to one another in their descriptions than were freshmen.

LSA was used as a tool for investigating the content of knowledge representations. This type of technique has its strengths and weaknesses. Its strengths are that it is an automated technique that can identify differences among writing samples, and it is consistent with at least one other type of analysis. However it is difficult to use as a purely exploratory tool due to the high dimensionality space in which documents and words are represented. Using such a high dimensional space means that questions must be formulated in terms of locations in this multidimensional space. For instance assessing the functional content of each document in this study required the creation of a vector that represented a document high in functional content. It may be difficult or impossible to formulate certain questions of interest in terms of locations in a multidimensional space. In cases where this is possible, LSA is able to answer questions in a more automated way than existing techniques. Another potential concern is that LSA may not be able to capture content across a set of words. For instance, a student may describe the functioning of a set of components across a phrase. LSA treats each document as just a group of words without considering order information. A human coder would be able to recognize the meaning of the phrase, but LSA may not be able to. We have attempted to alleviate this concern by showing that the LSA measure we used was well correlated with the results from human coders.

4. General discussion

The results from both studies support the idea that more experienced engineering students represent and reason about the functionality of a device and groups of components better than less experienced students. This is not to say that freshmen cannot or do not represent functional content, but rather that they are less able to construct or maintain such representations. The results are consistent with the kind of hierarchical chunking structure that has been proposed for knowledge representation in a number of domains. One such model has been presented in Fig. 7, which is loosely based on a description of conceptual chunking by Egan and Schwartz (1979). The main characteristics of this model are that devices are composed of functional units, which include groups of components acting to achieve some function. In this type of representation, function is the glue that holds everything together in a coherent representation.

Our results show that freshmen differ from seniors in their ability to represent the intermediate levels where chunks of components perform some subfunction of the device. The freshmen are able to chunk components together, but their chunks do not necessarily correspond to a set of meaningful functional units that together produce the behavior of the device. In this respect, seniors have a more integrative and less fragmented representation of the relation of the device and the components of which it is composed. This framework makes it easier for the seniors to recall devices because they have more constraints with which to reason. For instance, seniors can reason about which components go together to perform certain functions, and they can also reason about which functions may be necessary for overall device behavior. Freshmen on the other hand are less able to do this, and so they have more difficulty connecting together chunks and are less likely to talk about this abstract level of functionality when describing a device. This type of hierarchical representation also has implications for how devices are perceived and designed by engineers.

Observations of engineers have shown that they begin working on a design problem by decomposing it into modules based on the constraints of the problem with each module performing a certain function in the device, and it is rare for this decomposition to change during the design process (Ball et al., 1994). This process seems analogous to the way experts at solving physics problem can easily identify categories of problems and proceed to solve a problem with the appropriate schema (Chi et al., 1981). It is easy to see how the representation structure identified here supports this process of decomposing a device into separate functional modules. In fact if the top node of Fig. 7 represents a category of devices instead of one device, then this hierarchical structure allows the engineer to easily retrieve an appropriate plan for device decomposition from memory, given that a certain type of device is desired. Depending on how routine the design task is, this representation may then allow for the activation of appropriate schemas that enable the engineer to easily produce solutions that accomplish each module's function. This illustrates how using an appropriate knowledge representation can transform a design problem with a potentially intractable search space into a manageable problem.

The results provide further insight into the early stages of expertise acquisition, which have received relatively little attention. The transition from novice to expert in this domain appears to involve being able to identify meaningful groups of components that perform a function as well as being able to represent and reason about how these groups of components interact to produce the overall device behavior. Seniors have a more coherent representation of devices, and the main reason for this coherence seems to be the importance they place on function. By breaking the device into functional chunks, the device can be understood as a series of functions that produce some transformation from input to output. This same principle then applies to the chunks that were created, as each one can be broken up into a set of interacting components that performs some function. More complex devices than the ones studied here may require additional levels of functional units in traversing the levels between individual components and the overall device.

It seems likely that further changes in representation would occur as these engineers become professionals and acquire further experience. Although the work presented here focuses on the early acquisition of expertise and not on professionals, it does raise some interesting issues concerning the representations that experts might use and has potential implications for engineering education. Future work should include longitudinal studies that would delineate the time course of learning and help in identifying potential improvements in engineering education. Although learning is the most likely explanation for our findings, such studies would also completely rule out other less likely causes such as attrition and selection effects.

In general, the early stages of expertise appear to involve the creation of advanced knowledge structures that support expert performance in the domain. The knowledge representations appear to change from fragmented structures into more advanced coherent representations as expertise is acquired. This process can also be seen in the multiple representations of medical students who have assembled a network of causal scientific knowledge but have not yet linked it together with practical clinical knowledge (Hmelo, 1998; Patel, Evans, & Groen, 1989). In our studies, the seniors possessed a unified and connected representation of devices through their attention to function, but the freshmen lacked such a coherent representation.

The function of groups of components is not an obvious property of any single component or device, but it is a higher order property that allows the expert to work in the domain. Our results concerning the early acquisition of expertise in this domain make it apparent that the attention to function is enabling the seniors to form a coherent representation of devices. This attention to the functional properties and relations of groups of objects in a domain can be seen in a number of different domains of expertise. Work in chess has shown that attack and defense relations between chunks of pieces are important components of a chess master's representation (Chase & Simon, 1973). In computer programming the function of keywords and groups of these keywords are important aspects of an expert's representation (McKeithen et al., 1981), and work on understanding how people understand complex systems has shown that functional relations are present in the representations of experts but not in novices (Hmelo-Silver & Pfeffer, 2004). It seems likely that functional properties play a key role in the development of a coherent representation in other domains as well.

It is clear that there are changes in the representation of device information as expertise is acquired. In some ways these changes are consistent with what is known to happen in a number of other domains of expertise, but there will always be some elements that are specific to the domain under investigation. The specific representational structures that arise in these different fields are always partially determined by domain-specific properties. In the domain of engineering design, the necessity to design complex functioning devices leads to a hierarchical representation where the parts are related through functional relations. The hierarchical decomposition of functions is a necessary part of the representation because it may not be possible to consider the entire device even for moderately complex devices. This kind of partial decomposability is one approach people use to deal with complexity (Simon, 1969). Computer programming has a similar concern for decomposing complex programs into smaller functions, and it has been shown that expert programmers also have a hierarchical representation concerned with functional relations (McKeithen et al., 1981). Physics is concerned with explaining the behavior of objects, using just a few underlying principles, and it has been shown that one of the primary structures for experts in this domain are schemas that identify the corresponding physical principle to use for a particular problem (Chi et al., 1981). These are just a few examples, but the general point is that the characteristics and demands of the domain combined with the limitations of the cognitive system lead to the various representational structures seen in different domains of expertise.

Understanding these structures and how they arise provides important insights into the structure and processes of the mind, and it leads to a better understanding of performance in the domain, which can lead to improvements in cognitive aids as well as in education. For example, the different representations that the freshmen and seniors use may help explain differences in the design process that novices and experts use. As noted before, it has been shown that novices tend to approach design problems in a depth-first manner, whereas experts use a breadth-first approach (Ball et al., 1997). This could be due to a fragmented representation of the device as the novices have more trouble seeing how the various subsystems interact and decide to complete one subsystem before moving on to the next. If this is the case, then students may benefit from instruction that makes the functional properties of groups of components more concrete. This may help them incorporate important but abstract functional properties into their representations of devices. The studies reported here address differences in representation. Future work should focus on mechanisms that can explain how these changes occur and what specific information and experiences lead to these changes.

Acknowledgments

This research was sponsored by a National Defense Science and Engineering Graduate Fellowship and by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant number FA9620–04–1-0201. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of AFOSR or the U.S. Government.

References

Alsruhe, W. (2001). *U.S. Patent No. 6,467,556.* Washington, DC: U.S. Patent and Trademark Office.

Atman, C. J., Chimka, J. R., Bursic, K. M., & Nachtmann, H. L. (1999). A comparison of freshman and senior engineering design processes. *Design Studies, 20,* 131–152.

Baggett, P., & Ehrenfeucht, A. (1988). Conceptualizing in assembly tasks. *Human Factors, 30,* 269–284.

Ball, L. J., Evans, J. S. B. T., & Dennis, I. (1994). Cognitive processes in engineering design: A longitudinal study. *Ergonomics, 37,* 1753–1786.

Ball, L. J., Evans, J. S. B. T., Dennis, I., & Ormerod, T. C. (1997). Problem-solving strategies and expertise in engineering design. *Thinking and Reasoning, 3,* 247–270.

Ball, L. J., & Ormerod, T. C. (1995). Structured and opportunistic processing in design—A critical discussion. *International Journal of Human–Computer Studies, 43,* 131–151.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4,* 55–81.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121–152.

Deerwester, S., Dumais, S. T., Furnas, G. W., & Landauer, T. K. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41,* 391–407.

Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. *Memory and Cognition, 7,* 149–158.

Feltovich, P. J., Spiro, R. J., & Coulson, R. L. (1989). The nature of conceptual understanding in biomedicine: The deep structure of complex ideas and the development of misconceptions. In D. A. Evans & V. L. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling* (pp. 113–172). Cambridge, MA: MIT Press.

Gobet, F. (1998). Expert memory: A comparison of four theories. *Cognition, 66,* 115–152.

Gobet, F., & Simon, H. A. (2000). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science, 24,* 651–682.

Goel, V. (1995). *Sketches of thought.* Cambridge, MA: MIT Press.

Heiser, J., & Tversky, B. (2002). Diagrams and descriptions in acquiring complex systems. In W. D. Gray & C. D. Schunn (Eds.), *24th Annual Meeting of the Cognitive Science Society* (pp. 447–452). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hill, A., Song, S., Dong, A., & Agogino, A. (2001, September). *Identifying shared understanding in design using document analysis* (Report No. DETC2001/DTM-21713). Paper presented at the 13th International Conference on Design Theory and Methodology, Pittsburgh, PA.

Hmelo, C. E. (1998). Problem-based learning: Effects on the early acquisition of cognitive skill in medicine. *Journal of the Learning Sciences, 7,* 173–208.

Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science, 28,* 127–138.

Koedinger, K. R., & Anderson, J. R. (1990). Abstract planning and perceptual chunks—Elements of expertise in geometry. *Cognitive Science, 14,* 511–550.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211–240.

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science, 208,* 1335–1342.

McKeithen, K. B., Reitman, J. S., Rueter, H. H., & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology, 13,* 307–325.

Moss, J., Cagan, J., & Kotovsky, K. (2004). Learning from design experience in an agent-based design system. *Research in Engineering Design, 15,* 77–92.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Oxford, England: Prentice Hall.

Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987, October 30). Teaching reasoning. *Science, 238,* 625–631.

Patel, V. L., Evans, D. A., & Groen, G. J. (1989). Biomedical knowledge and clinical reasoning. In D. A. Evans & V. L. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling* (pp. 53–112). Cambridge, MA: MIT Press.

Quesada, J. F., Kintsch, W., & Gomez, E. (2002). A theory of complex problem solving using latent semantic analysis. In W. D. Gray & C. D. Schunn (Eds.), *24th Annual Conference of the Cognitive Science Society* (pp. 750–755). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Reitman, J. S. (1976). Skilled perception in Go: Deducing memory structures from inter-response times. *Cognitive Psychology, 8,* 336–356.

Reitman, J. S., & Rueter, H. H. (1980). Organization revealed by recall orders and confirmed by pauses. *Cognitive Psychology, 12,* 554–581.

Richman, H. B., Staszewski, J. J., & Simon, H. A. (1995). Simulation of expert memory using epam iv. *Psychological Review, 102,* 305–330.

Simon, H. A. (1969). *The sciences of the artificial.* Cambridge, MA: MIT Press.

Smith, R. P., & Leong, A. (1998). An observational study of design team process: A comparison of student and professional engineers. *Journal of Mechanical Design, 120,* 636–642.

Stone, R. B., & Wood, K. L. (2000). Development of a functional basis for design. *Journal of Mechanical Design, 122,* 359–370.

Wood, K. L., & Greer, J. L. (2001). Function-based synthesis methods in engineering design. In E. K. Antonsson & J. Cagan (Eds.), *Formal engineering design synthesis* (pp. 170–227). Cambridge, England: Cambridge University Press.