COGNITIVE INVESTIGATIONS INTO KNOWLEDGE REPRESENTATION IN ENGINEERING DESIGN

Jarrod Moss, Kenneth Kotovsky
Department of Psychology, Carnegie Mellon University, USA

Jonathan Cagan
Department of Mechanical Engineering, Carnegie Mellon University, USA

Abstract. As engineering students 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 uses a recall paradigm, and the second uses Latent Semantic Analysis (LSA) to analyze brief descriptions written by engineering students. Both studies find that the most prominent differences between these two groups of students are their representations of the function of electromechanical components and how these components interact. The findings from these studies highlight some ways in which the structure and content of mental representations of design knowledge differ with experience.

1. Introduction

Engineering design is a domain in which a number of complex problem solving activities occur. As in all such tasks, cognitive processes operate upon the internal representations of the task as well as upon other relevant knowledge. These representations can change over the course of experience in order to enable a person to better respond to the problems and challenges of a domain. These representation changes are a reflection of the structure and content of a domain as well as the cognitive learning mechanisms responsible for the changes.

One motivation for studying expertise is to learn more about the general cognitive mechanisms which allow people to become experts in some domain given sufficient learning and practice. Eventually a person acquires both knowledge structures and cognitive processes that are specific to the domain of expertise and allow for efficient functioning within that domain. The knowledge structures and processes of expertise that arise are a function of both general cognitive mechanisms and the actual structure and content of the domain in which expertise is being acquired. This means that the content and to some extent the structure of the knowledge representation are determined by the particular domain of expertise. The structure of experts' representations across different domains may show some similarities either because they are constructed by the same learning mechanisms or because the domains of expertise have some common characteristics.

There are at least two types of reasons to study expertise in a particular domain. The first is to learn about the specific mental representations and cognitive processes employed by experts in that domain. This information is potentially beneficial in the design of cognitive aids that can assist experts or in improving the education of future experts in that domain. The second type of reason to study expertise in a domain is that it provides further insights into the cognitive learning mechanisms that produce the changes in representation and cognitive processes seen in expertise acquisition. One way to study these learning mechanisms is to see how they interact with the structure and content of a variety of domains in order to produce the mental structures and processes seen in experts across these domains.

A number of domains of expertise have been studied, and there have been some general findings about how the structure of domain knowledge changes with experience in these tasks. For instance, in games like chess and Go, a hierarchical database of commonly occurring piece configurations appears to exist in experts but not in novices (Chase and Simon 1973, Reitman 1976). This knowledge aids the expert in classifying the current situation and identifying good moves. Similar types of hierarchical chunking have also been identified in electronics technicians (Egan and Schwartz 1979). One of the general findings in these and other areas of expertise is that a hierarchical knowledge structure is often a component of expertise. Other general findings in the expertise literature are that experts tend to work forward from the givens in the problem rather than backwards 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 et al. 1981, Larkin et al. 1980).

Experts' knowledge representations enable them to handle problems and process information differently than novices. For example, physics experts solve physics problems in a different manner than do novices (Larkin et al. 1980). Physics experts can recognize and solve common problems in a more efficient manner than novices. 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 & Simon (1973) argue that what they see is perceptual chunking, and elsewhere it has been shown that while domain knowledge does not directly transfer to other domains it can still influence the way people reason about more general situations outside of their domain of expertise (Nisbett et al. 1987). So 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.

These effects of expertise all relate to the ability of the expert to process more domain information in a fixed period of time than can a novice. Experts have highly organized memory structures like schemas, templates, and retrieval structures (Gobet 1998, Richman et al. 1995). These memory structures allow for the easy retrieval and storage of domain information, and they affect how domain information is perceived. As information about a new problem is perceived, this information automatically activates relevant domain knowledge and processes. This allows experts to easily recognize and categorize information and solution schemas in their domain. Parts of the problem solving process are therefore more automatized in experts than they are in novices, and this enables experts to solve problems in a more efficient manner. In order to understand how experts in a domain solve problems it is necessary to examine the way domain information is represented.

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. As discussed below, there are a few studies that have examined cognition and expertise in engineering design through verbal protocols and other methods, but these studies usually deal with cognition at a coarse level and do not examine mental representation. The work presented here is an initial step towards a detailed examination of the representations and processes that allow engineers to perform the complex tasks required by their profession. This paper presents two studies which begin to answer the question of what kinds of representation changes accompany the transition to expertise.

In particular, the studies presented here look at freshmen and senior engineering students in order to see what kind of representation changes accompany the early transition to expertise. The differences between freshmen and seniors may generalize to professional engineers upon further investigation, or alternatively the transition from student to professional may involve other qualitative changes in mental representation. Two different methodologies were utilized in examining representation differences in the two groups of students. The first study utilizes a recall paradigm that has been employed by a number of researchers looking at expert/novice differences (e.g., Chase and Simon 1973). This first study examines some basic differences in how components in devices are represented and chunked together. The second study uses Latent Semantic Analysis (Deerwester et al. 1990) as a methodological tool to aid in exploring and analyzing the content of students' representations. This study seeks to determine whether the seniors think about and represent devices in a more abstract functional manner than do freshmen.

2. Expertise in Design

There has been some relevant work on the differences between experts and novices in engineering design. For instance Atman, Chimka, Bursic, and Nachtmann (1999) have looked at differences in the design processes of freshmen and senior students by analyzing concurrent verbal protocols. They found that seniors have a better representation of what a good design process entails and can transition between steps in the design process more easily than can freshmen. Also, seniors 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 the differences between the design processes of students as compared to 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 and 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 using concurrent verbal protocols (Ball et al. 1997, Ball et al. 1994, Ball and Ormerod 1995). These studies found that novices use a depth-first design process while experts use more of a breadth-first approach. Both groups of designers decomposed the problem into modules, but experts tended to develop each module to a certain level of detail before moving to the next level of detail. Novices were more likely to do detailed design on one module before moving on to the next. It is proposed that the depth-first structure is advantageous for novices since it limits the amount of goal information they must store in memory. Even when experts deviated from their breadth-first structure there seemed to be principled reasons for doing so. For example, an expert may quickly follow one potential solution in depth to assess its feasibility before proceeding to other parts of the design.

There has also been some work in the domain of architectural design. For instance, Suwa and Tversky (1997) collected a set of retrospective protocols from students and a couple of professional architects. They found that the experts tended to follow certain trains of thought in more depth than the novices. Also the experts were better at reading certain types of functional information from their sketches than were novices. This same set of protocols has been the basis for other work as well (Kavakli and Gero 2001, Kavakli and Gero 2002, Suwa et al. 1998), but these have all been case studies in which either a single expert or one expert and one student have been examined so it is impossible to determine if there are any statistically reliable differences in these studies. These studies have all used retrospective protocols because it was believed that concurrent verbalization would bias the design and sketching behavior of participants. The retrospective protocols done in these studies used a video of the design session to cue verbalization. This procedure is likely to bias the results since the only recall cues to the designer are from sketching behavior. Other converging methods should be used to confirm these results to make sure the results generalize to other groups and to insure results were not overly biased by the retrospective method used. This is just one set of examples of work in the domain of architectural design that may relate to expertise in engineering design. Since it is unlikely that specific results in this domain will transfer directly to engineering design, there is a need to analyze the content and structure of the two domains in order to determine which results are relevant to expertise in engineering design. As discussed above, the cognitive learning mechanisms that enable expertise acquisition are the only thing guaranteed to hold across domains, and most studies of expertise in design do not study cognition at this level.

Overall, the work on engineering design and related areas tends to focus more on differences in the design processes of experts and novices than on representation and other cognitive issues. 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) presents an analysis of design problem spaces and the results of a study which indicate that certain types of symbol systems are necessary to support design activity. His work is concerned with some general necessary properties of a representational system. However, the subjects in this study were all experts so it is unclear how these results map onto novices. As stated above, it is necessary to understand how devices and other domain knowledge are represented in memory and how these representations change with expertise. The studies presented below are a first step in understanding these important issues.

3. Experiment 1: Chunking of Components

This study examines how the participants chunk components into larger meaningful units. Just as experts are known to chunk elements into larger units of knowledge in other domains such as electronics (Egan and Schwartz 1979), 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 which 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. However, engineers will also be able to reason about the functionality of a particular device and break it into functional units regardless of whether they are commonly occurring or not.

In order to investigate these issues, a recall paradigm was utilized that extends an approach used by others to study chunking differences in expert/novice behavior (Chase and Simon 1973, Reitman 1976). The basic method is to present a stimulus, such as a chess board in a mid-game 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 inter-response times (IRTs) 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 a problem in the chess research since the chunks in chess have more of a hierarchical relationship. In our experiment, only a recall task was used, and in order to avoid problems with finding an appropriate IRT boundary, analysis of IRTs was only one of many measures used to look at representation differences. In particular, we looked at percent recall after one exposure, errors, patterns of recall, and alternate methods of identifying chunks in addition to IRTs.

3.1. METHOD

3.1.1. Participants

Fifteen seniors majoring in mechanical engineering volunteered for the study. These students were recruited from a required senior engineering design course at Carnegie Mellon. Fifteen freshmen engineering students 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.

3.1.2. Stimuli

Three electromechanical devices were represented in schematic diagrams which indicated how components fit together in each device. The schematics were represented in an idealized fashion where only the types of components and the connections between these components were displayed. For example, all gears were represented by the same icon which includes no information about different sizes, shapes, or types of gears. Connections between components were represented by lines connecting components. An example design schematic is shown in Figure 1. The number of components in each device was 16, 13, and 14 for the drill, pressure gauge, and weighing machine respectively. The number of connections in each device is 9, 11, and 12 for the drill, pressure gauge, and weighing machine respectively. The weighing machine is similar in purpose to a bathroom scale. The number of unique components differs since some types of components were used more than once in a design. The drill had only 9 unique components, while the pressure gauge and weighing machine had 11 and 12 unique components respectively. Each diagram also had a label at the top indicating the type of device depicted as shown in Figure 1.

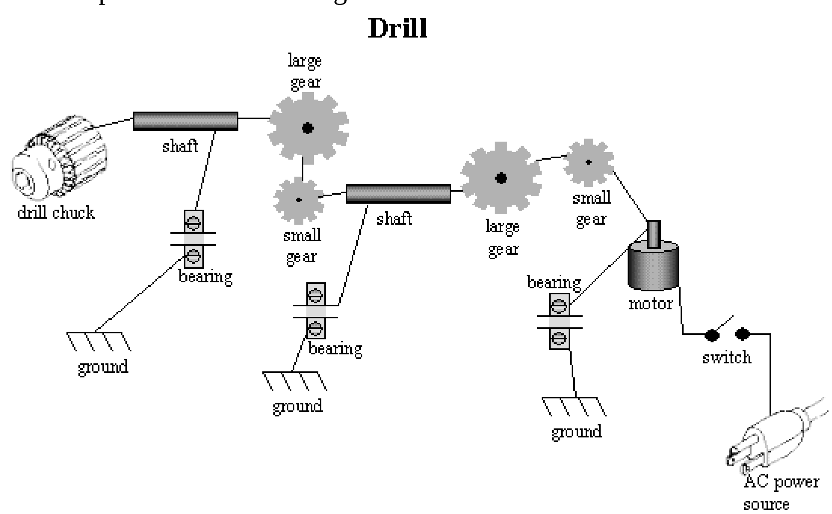

Figure 1. Example of diagrams seen by participants

3.1.3. Procedure

Participants were asked to recall three design schematics using a graphical interface after a brief study period. Participants received instruction and were allowed to become familiar with the interface and the type of representation used in the diagrams. The user interface is depicted in Figure 2, 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 seconds, and then the display of the user interface replaced the schematic. Participants then had 3 minutes in which to recall as much of the schematic as possible. The design schematic was presented again for 40 seconds if the participants had not recalled the design completely. These periods of display and recall alternated until the participant recalled the device perfectly. 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 for purposes of analysis.

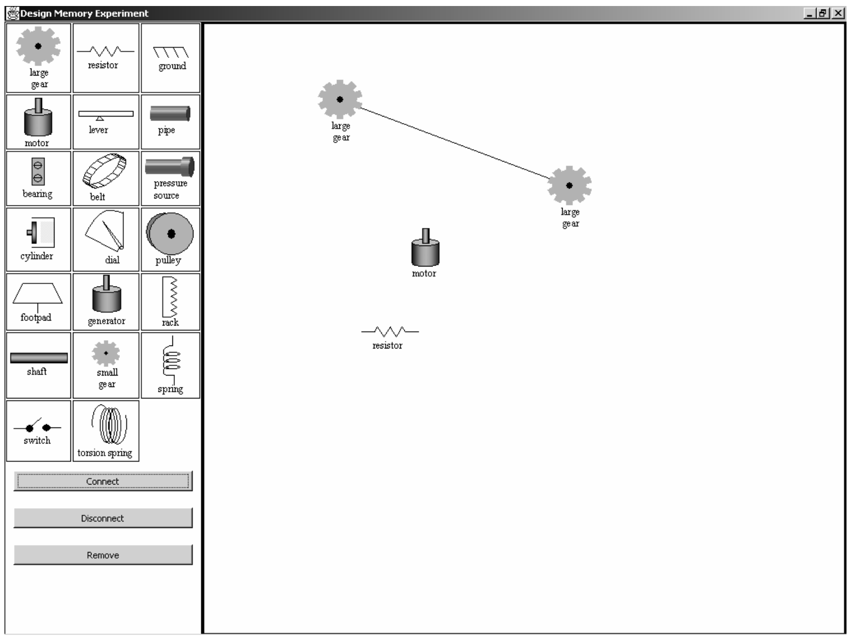

Figure 2. Screenshot of user interface

3.2. RESULTS

The percentage of components and connections between components recalled correctly during the first recall session of a trial was analyzed. Recall of components was almost perfect for all devices for both freshmen (M = 92.4%, SD = 9.23%) and seniors (M = 93.6%, SD = 10.2%), and experience level had no significant effect. Device type does have an effect on this measure, $F(2,56) = 3.63$, $p = .03$, and further contrasts showed that the drill components (M = 95.8%, SD = 7.58%) were recalled significantly better than both the weighing machine (M = 91.2%, SD = 10.0%), $F(1,28) = 7.81$, $p = .01$, and pressure gauge components (M = 92.3%, SD = 10.3%), $F(1,28) = 5.05$, $p = .03$. Recall of connections was lower overall than for components. There was no significant difference between freshmen (M =

77.8%, SD = 20.6%) and seniors (M = 82.8%, SD = 23.6%) on recall of connections, but again there was a significant effect of device type, $F(2,56) = 5.96$, $p = .005$. The drill's connections (M = 89.3%, SD = 14.6%) were recalled better than both the weighing machine (M = 73.8%, SD = 25.9%), $F(1,28) = 11.4$, $p = .002$, and pressure gauge (M = 78.9%, SD = 21.5%), $F(1,28) = 6.83$, $p = .014$. These results indicate that some devices were harder to recall than others. The most difficult design to recall seems to be the weighing machine followed by the pressure gauge, and the easiest to recall is the drill. The drill has only 9 unique components with a number of repeating patterns, but the other two devices have an increasing number of unique components. It makes sense that recall difficulty would increase with the more components and locations that have to be remembered.

A number of error types were defined and analyzed. They include adding a component that is not part of the design, removing a component that is needed, connecting two components that are not connected in the original design, and disconnecting two components that should be connected. Freshmen made more errors overall than seniors, $F(1,28) = 42.1$, $p < .001$. Device type did not have an effect on overall errors, and there was no interaction between device type and experience. The component removal and disconnect errors did not occur frequently enough to be analyzed separately, but analyses were done on the add and connect errors. There were no effects of device type or experience on the add errors, but there was an effect of experience on the number of connection errors, $F(1,28) = 43.3$, $p < .001$. Connection errors can be further divided into possible connections and impossible connections depending on whether the two components could actually be connected in the real world. Connection errors are displayed in Figure 3, where it can be seen that freshmen do make more connection errors than seniors. There was no significant interaction between device type and experience with respect to connection errors.

Patterns of reconstruction were also analyzed, and one meaningful pattern was identified. 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 Figure 1, 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. 14 out of the 15 seniors used this pattern at least once, while only 8 of the 15 freshmen did. In addition, two of the three devices were presented so that their input was on the left and power moved through the device from left to right, but the drill was presented so that its input was on the right side of the screen. For the drill, 6 seniors and only 2 freshmen went from input to output. One of these seniors and one of these freshmen actually saw the drill as the first design, but in the other cases the participant had already reconstructed another device from left to right (input to output) and reversed for the drill. This provides some evidence that seniors prefer to reconstruct the device based on the flow of energy through the components of the device, and there was some preference for moving from input to output even when the direction of input to output was reversed on the display.

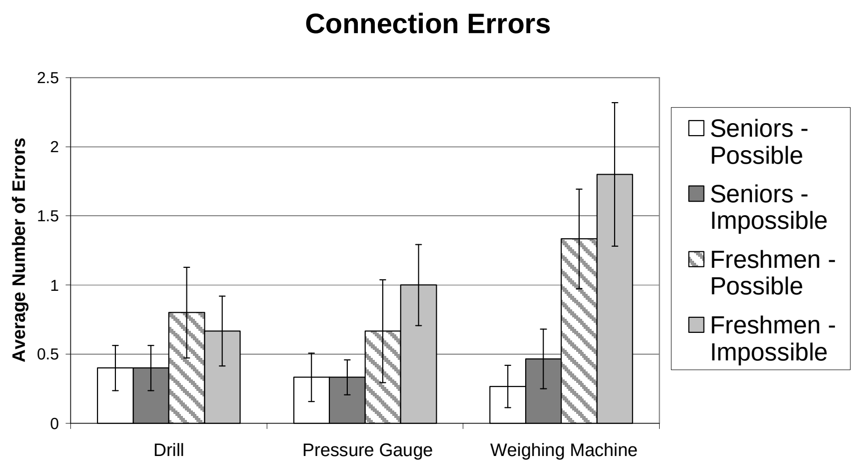

Figure 3: Average number of connection errors

The data can also be divided into chunks, but these divisions were not based solely on IRTs. First, an IRT criterion was set to distinguish between chunk transitions from within chunk transitions. A cutoff of 4 seconds was used. This value may be conservative as other studies have used boundaries of around 2 seconds. However, without an additional task such as the perception task used by Chase & Simon (1973) it is difficult to come up with a definite boundary. For example, in the perception task participants had to reconstruct a mid-game chess board, but they were able to glance back and forth from the board to be reconstructed and the board on which they were reconstructing the game position. It was assumed that participants encoded one chunk during each glance. Between chunk IRTs can then be distinguished from within chunk IRTs by looking at the difference in IRTs for when a participant recalled two pieces in succession without a glance at the original board and when they recalled two pieces separated by a glance at the original board.

The 4 second cutoff only includes 20% of the chunk transitions in our study. So the vast majority of transitions are still classified as within chunk transitions. However, the structure of the task also defined additional chunk boundaries. In the process of reconstructing a device, many participants would add a set of components and then proceed to connect those components before adding another set of components. This seems to provide a natural boundary whereby a participant adds a chunk, connects the components in it, and then adds another chunk. Using these two types of chunk boundaries, the data was segmented into individual chunks. Participants also drew chunk boundaries as mentioned before so their drawn chunks could be compared to the chunks generated from the recall data.

As a first step, the number of times a specific set of components were chunked together by participants was calculated. This allowed assessment of the chunks that individuals agreed on to some extent. The chunks identified by analysis of the recall data were similar to the chunks identified by the participants. This type of group level analysis indicates that both freshmen and seniors agree on the same types of chunks with the seniors being somewhat more consistent in the chunks they identify. Freshmen and seniors both produce chunks having between 2-3 components per chunk. While there seems to be agreement between freshmen and seniors when it comes to what should be chunked, the chunks identified do explain the error results mentioned earlier as will be explained below.

As mentioned before one of the most common types of errors was the connection errors, and this was the only type of error where the frequency of the error differed for freshmen and seniors. For both freshmen and seniors, 95% of their errors occurred when connecting two components that were not in the same chunk. When making a connection error, both groups are likely to make the error when connecting two different chunks, but freshmen make 2-3 times more of these errors than seniors depending on the particular problem. The difference in the frequency of connection errors then reflects the ability of seniors but not freshmen to remember how chunks of components were connected together.

3.3. DISCUSSION

In general it appears that seniors differ from freshmen on their understanding and ability to remember information about the connections and interactions between components. Seniors make fewer errors than freshmen, and the analyses indicate that this is mostly due to increased connection errors for freshmen. There is also some indication that freshmen may make more connection errors as problem difficulty increases (Figure 3), but this interaction failed to reach significance probably due to lack of statistical power. Seniors tend to rely more on recall methods that utilize the natural flow of power from one component to the next than do freshmen. From the connection error and chunking results, it is apparent that freshmen have more difficulty remembering how chunks of components connect together. Therefore, one of the main differences between the groups appears to be in their ability to remember how chunks of components connect together to form the overall device. This implies that freshmen are able to chunk components but have more difficulty connecting these functional units together to produce overall device behavior. The representation of chunks of components is weaker in freshmen than in seniors. While seniors are able to remember the chunks and how they interact to produce overall device behavior, freshmen are not as able to represent such interactions.

4. Experiment II: Functional Reasoning

In order to determine if a kind of abstract functional understanding was more prevalent in the mental representations of seniors than freshman a second study was run. Since it is apparent from the first study that freshmen lack a strong representation of how components interact, this should mean that freshmen are not as able to reason about devices in an abstract functional manner since the abstract functional level would entail reasoning about the functions of chunks of components and how those functions interact in producing device behavior.

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 and Tversky 2002). In the current study it is assumed that the participants are all high in mechanical ability since they are all majoring in or intending to major in mechanical engineering. Based on these results and those of Experiment I it is hypothesized that more experienced participants will demonstrate better utilization of functional knowledge than less experienced participants. In order to investigate these issues, a new method of data analysis will be introduced. In the Heiser and Tversky (2002) work, 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) will be used to test for higher functional content in written text. This method does not require someone to decide whether each proposition contains functional information or not. The data could also be analyzed using the proposition coding system as well just to show that the two methods are consistent, and this is part of planned future work.

In order to investigate these issues, students were asked to write brief descriptions of devices that were presented in diagrams. One assumption underlying this study 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 above, another hypothesis that will be investigated is that seniors may be 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 all to think about the devices in a similar manner.

4.1. LATENT SEMANTIC ANALYSIS

The participants' descriptions were analyzed using Latent Semantic Analysis (LSA). LSA was originally developed as an information retrieval technique designed to overcome synonymy problems (Deerwester et al. 1990). It has also been used for a number of other purposes including as a model of text comprehension (Landauer and Dumais 1997). More recently it has also been used to develop similarity metrics to be utilized in the analysis of data from complex problem solving trials (Quesada et al. 2002). LSA begins with a term-by-document frequency matrix, and produces a reduced dimensionality space in which each document or term can be seen as a point or vector in that space. Similarities can then be computed between any two terms or documents by computing the cosine between the appropriate vectors. These properties make LSA an excellent tool for exploring similarities and differences between documents written by participants, thus shedding light on the content of their representations.

One set of researchers has already utilized LSA to try to identify shared design understanding among a set of designers (Hill et al. 2001). That study was concerned with building information management tools that retrieved relevant information based on a certain shared understanding of the design problem. This shared understanding was determined by using LSA to analyze documentation from design projects. This use of LSA allows for a particular- representation of the desired design project to be created, but these researchers were not concerned with identifying properties of this representation.

4.2. METHOD

4.2.1. Participants

In this study, there were 44 volunteers from a senior mechanical engineering design class. There were also 24 freshmen volunteers from a freshman mechanical engineering class, and the study was run during their first mechanical engineering course.

4.2.2. Stimuli

Three electromechanical device diagrams were used in this study. These diagrams were taken from patents for a power screwdriver (Figure 4), 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 labels were removed in order to ensure a similar number of labeled components for each diagram. Each diagram had 9-11 labeled components and had the name of the device printed in large bold letters at the top.

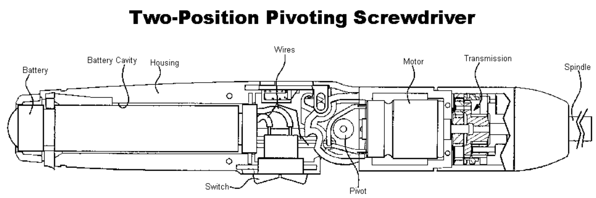

Figure 4. Example diagram seen by participants

4.2.3. 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 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 the current state of their description was saved to a time stamped file. Participants were instructed to spend about five minutes describing each device. They were not forced to spend exactly five minutes on a device, but they had to pace themselves to finish all three descriptions in 18 minutes. There was a clock displayed in the lower right corner of the screen to help them pace themselves.

The participants were then asked to rate their prior knowledge of each device (1=poor, 7=good). The freshmen participants then completed an additional set of eight true/false questions for each device before they gave their ratings. The questions consisted of a mix of questions emphasizing 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 during these questions. The freshmen all participated during the second half of their first semester. The seniors were divided into two groups of 20. The first group completed the study in the first three weeks of the semester, and the second group participated after half of the semester had passed. This timing was used because it allowed us to test the hypothesized that the senior design course may be instrumental in changing the way students thought about designs since it specifically included a lecture on function structures about four weeks into the course.

4.3. RESULTS

Four seniors and two freshmen were excluded from all analyses since they failed to finish in the allotted time. This included two seniors from the early group and two from the late group.

Freshmen rated themselves as having more knowledge about the weed trimmer than the other two devices ($\chi^2 = 6.87$, $p = .03$), but there were no significant differences between devices in the seniors' ratings. There were also no significant differences between the ratings of freshmen and seniors for a particular device. All groups were therefore similar in their prior knowledge of the devices. The true/false questions were included to assess functional and structural knowledge of 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 somewhere between 50 and 300. Most other LSA work has been done with much larger text corpora and optimal dimensionality was around 300 dimensions. Since our corpus of device descriptions is much smaller, a smaller number of dimensions are 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. A separate average was produced for freshmen and seniors for each device. For example, all freshmen drum 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. This vector was then treated as a line in the multidimensional space with its origin at the average freshman description.

All descriptions for a particular device were then orthogonally projected to a location on this line. Their location provides a way of examining freshmen and senior differences.

Due to the way the line is constructed and the way documents are projected to points on this line, the average senior description for a device will be a certain distance away from the average freshman description. The first test is whether there is a significant difference between where the senior and freshmen descriptions fall on this line, and they are significantly different, $F(1,54) = 121$, $p < .001$. The earlier hypothesis that seniors will be more similar to each other as a group than freshmen was then tested by looking at how far freshmen and seniors are from the average freshman and senior descriptions. For instance, seniors should on average be closer to the average senior description than freshmen are to the average freshman description. This difference is also significant with seniors deviating less from their average description than freshmen, $F(1,54) = 275$, $p < .001$. This means that seniors are more consistent with each other than freshmen on the information they include in descriptions of a particular device.

The search engine qualities of LSA were utilized in order to examine the hypothesis that seniors included more information about the functioning 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. In order 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) have developed a vocabulary to explain the internal chain of functions that produces a device's behavior. They have shown 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. 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.

This process functions like a search engine, and the system ranks documents according to their similarity to this query consisting of function words. Both experience level, $F(2,61) = 3.7$, $p = .03$, and device type, $F(2,122) = 12.7$, $p < .001$, have significant effects on a document's similarity to this query, but these two factors did not interact. Further contrasts reveal that the freshmen have significantly lower cosines (i.e., are less similar to the query) than the later group of seniors, $F(1,61) = 5.05$, $p = .02$. Also the drum brake system descriptions had higher cosines than both the screwdriver, $F(1,61) = 18.4$, $p < .001$, and the weed trimmer, $F(1,61) = 16.7$, $p < .001$.

Document relatedness rankings retrieved from a query are often easier to interpret than the cosine values. Rankings are determined by sorting the cosines between a document and the query in descending order. The document with the highest cosine is ranked 1 and so on. The average rank (out of 192) for freshmen descriptions was 110.2 (SD = 55.4), while the seniors on average ranked 94.6 (SD = 55.5) and 82.0 (SD = 52.7) for the earlier and later groups respectively. These results indicate that seniors included more content that is similar in meaning to the function words in the query. Furthermore, the earlier group of seniors is ranked between the freshmen and the later group of seniors indicating an increase in functional content with experience.

4.4. Discussion

The results from this study agree with and support those of the first study in that seniors are shown to incorporate more function information into their representations. Seniors do differ from freshmen on their similarity to the prototypical or average descriptions, and they differ on the amount of functional content they include in their descriptions. This means that seniors have all adopted a similar representation of the device, and that this representation includes more functional content than do the representations that freshmen use. The fact that seniors include more function content in their description adds further support to the idea that one of the main differences between the two groups of students is the ability to represent and process the functionality of chunks of components in a device.

5. 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 its components better than less experienced students due to differences in the representations used by the two groups. This finding seems to be the main difference in design knowledge representation at these levels of experience. This is not to say that freshmen can not or do not represent functional content, but instead that the memory structures that support this type of reasoning are not as well developed in freshmen. This level of representation provides the seniors with additional constraints when recalling the devices presented in the first study.

Senior engineering students may have a more detailed network of design knowledge which integrates content which the freshmen hold in separate representations if at all. For instance, both a freshman and senior may have similar representations of the structure of a gear and how the gear interfaces with other components. However, the senior may also have a representation of the abstract function of a gear and a set of contexts in which a gear may perform well. This idea of context is important because some functions can be performed by a group of two or more components, but not by any one of the components by itself. For instance, transforming rotational motion into translational can be performed by a rack and gear together but not by either of them separately. It may be that building up a set of such associations between sets of components and functionality is one of the main changes that take place as an engineer gains more experience. This type of learning and representation change seems similar to the learning of chunks in chess and other domains (Chase and Simon 1973).

One explanation for the findings in this paper is that the device is represented at multiple levels. The most abstract level is the overall function of the device, and the most detailed level would be the components which make up the device. In between these two levels are one or more levels in which the function chunks of components are represented. There may be multiple levels of this type of chunking in which the chunks at one level are grouped into sets at the next level. At any particular level, a chunk of components performs a subfunction which contributes to the overall functioning of a device. A similar model called *conceptual chunking* has been proposed to deal with the expertise of electronics technicians (Egan and Schwartz 1979). Using this type of model, it is proposed that the freshman differ from seniors in their ability to represent the middle levels where chunks of components perform some function. The freshmen seem to understand what types of components should go together to perform a sensible function, but they have problems linking these functions together to achieve the overall device function. In this view, 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 since they have more constraints from 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 may not be as able to reason about which functions are necessary for the overall behavior of the device, and so they have more difficulty connecting together chunks of the device and are less likely to talk about this level of function when describing a device.

This type of conceptual chunking relationship also relates to earlier work on the observed structure of the engineering design process in novices and experts. It has been found that experts adopt a more breadth-first design process, while novices use a depth-first process (Ball et al. 1997, Ball et al. 1994, Ball and Ormerod 1995). The designers in these studies all decomposed their design into more manageable modules. The observed structure of the design process could relate to how well developed the mental representations of the device are for experts and novices. Depth-first design processes require maintaining less intermediate information about other parts of a device, while a breadth-first approach requires thinking about how parts of a device interact at a number of different levels of specificity. It may therefore be easier for a novice to use a depth-first approach toward design. In this way, the work on representation presented here provides some insights into why design processes differ for novices and experts.

LSA is a potentially powerful tool for investigating the structure of knowledge representations. A number of interesting questions about representation can be answered by using this automatic technique to represent a set of documents in a multidimensional space. One of the main problems in using LSA as an exploratory tool is trying to find the correct number of dimensions. However, varying the dimensionality of the space could also vary the amount of detail incorporated in the representations being examined. Seen in this light, having a variable number of dimensions could be a positive aspect of LSA as an analysis tool since the amount of representational detail being examined could be varied with one parameter.

One limitation of this work is that it only deals with engineering students. There are plans to expand this work to professional engineers, and it should be interesting to see how even greater amounts of design experience affect a person's representation of design knowledge. However, this work does capture some of the differences in the beginning stages of the acquisition of expertise. Also, even though differences associated with design experience have been identified, there is currently no mechanism that explains how these changes come about. Generating such an explanation is a necessary step in coming up with a cognitive model of the engineering design process.

Acknowledgments

This research effort 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 F49620-01-1-0050. 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

Atman, CJ, Chimka, JR, Bursic, KM and Nachtmann, HL: 1999, A comparison of freshman and senior engineering design processes, *Design Studies,* **20**(131-152.

Ball, LJ, Evans, JSBT, Dennis, I and Ormerod, TC: 1997, Problem-solving strategies and expertise in engineering design, *Thinking & Reasoning,* **3**(4)**:** 247-270.

Ball, LJ, Evans, JST and Dennis, I: 1994, Cognitive Processes in Engineering Design: a Longitudinal Study, *Ergonomics,* **37**(11)**:** 1753-1786.

Ball, LJ and Ormerod, TC: 1995, Structured and Opportunistic Processing in Design - a Critical Discussion, *International Journal of Human-Computer Studies*, **43**(1)**:** 131-151.

Chase, WG and Simon, HA: 1973, Perception in Chess, *Cognitive Psychology,* **4**(1)**:** 55-81.

Chi, MTH, Feltovich, PJ and Glaser, R: 1981, Categorization and Representation of Physics Problems by Experts and Novices, *Cognitive Science,* **5**(2)**:** 121-152.

Deerwester, S, Dumais, ST, Furnas, GW and Landauer, TK: 1990, Indexing by Latent Semantic Analysis, *Journal of the American Society for Information Science,* **41**(6)**:** 391-407.

Egan, DE and Schwartz, BJ: 1979, Chunking in Recall of Symbolic Drawings, *Memory & Cognition,* **7**(2)**:** 149-158.

Gobet, F: 1998, Expert memory: a comparison of four theories, *Cognition,* **66**(2)**:** 115-152.

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

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

Hill, A, Song, S, Dong, A and Agogino, A: 2001, Identifying shared understanding in design using document analysis, *13th International Conference on Design Theory and Methodology*, Pittsburgh, PA, DETC2001/DTM-21713.

Kavakli, M and Gero, J: 2001, Sketching as mental imagery processing, *Design Studies,* **22**(4)**:** 347-364.

Kavakli, M and Gero, J: 2002, The structure of concurrent cognitive actions: A case study on novice and expert designers, *Design Studies,* **23**(1)**:** 25-40.

Landauer, TK and Dumais, ST: 1997, A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge, *Psychological Review,* **104**(2)**:** 211-240.

Larkin, J, McDermott, J, Simon, DP and Simon, HA: 1980, Expert and Novice Performance in Solving Physics Problems, *Science,* **208**(4450)**:** 1335-1342.

Nisbett, RE, Fong, GT, Lehman, DR and Cheng, PW: 1987, Teaching Reasoning, *Science,* **238**(4827)**:** 625-631.

Quesada, JF, Kintsch, W and Gomez, E: 2002,A theory of complex problem solving using Latent Semantic Analysis In *24th Annual Conference of the Cognitive Science Society*(Eds, Gray, W D and Schunn, C D) Lawrence Erlbaum Associates, Mahwah, NJ, pp. 750-755.

Reitman, JS: 1976, Skilled Perception in Go: Deducing Memory Structures from Inter-Response Times, *Cognitive Psychology,* **8**(3)**:** 336-356.

Richman, HB, Staszewski, JJ and Simon, HA: 1995, Simulation of expert memory using EPAM IV, *Psychological Review,* **102**(2)**:** 305-330.

Smith, RP and Leong, A: 1998, An observational study of design team process: A comparison of student and professional engineers, *Journal of Mechanical Design,* **120**(4)**:** 636-642.

Stone, RB and Wood, KL: 2000, Development of a functional basis for design, *Journal of Mechanical Design,* **122**(4)**:** 359-370.

Suwa, M, Purcell, T and Gero, J: 1998, Macroscopic analysis of design processes based on a scheme for coding designers' cognitive actions, *Design Studies,* **19**(4)**:** 455-483.

Suwa, M and Tversky, B: 1997, What do architects and students perceive in their design sketches? A protocol analysis, *Design Studies,* **18**(4)**:** 385-403.