Cognitive Transformation Theory: 
Contrasting Cognitive and Behavioral Learning

Gary Klein, Ph.D., Holly C. Baxter, Ph.D. 
Fairborn, OH 
gary@decisionmaking.com, holly@decisionmaking.com

ABSTRACT

The traditional approach to learning is to define the objectives (the gap between the knowledge a person has and the knowledge the person needs to perform the task), establish the regimen for practice, and provide feedback. Learning procedures and factual data is seen as adding more information and skills to the person’s storehouse of knowledge. However, this storehouse metaphor is poorly suited for cognitive skill, and does not address the differing learning needs of novices and experts. Teaching cognitive skills requires the diagnosis of the problem in terms of flaws in existing mental models, not gaps in knowledge. It requires learning objectives that are linked to the person’s current mental models. It requires practice regimens that may have to result in “unlearning” that enables the person to abandon the current, flawed mental models. It requires feedback that promotes sensemaking. We propose a Cognitive Transformation Theory to guide the development of cognitive skills. Finally, we present several strategies that might be useful in overcoming barriers to understanding and to revising mental models.

ABOUT THE AUTHORS

Gary Klein, Ph.D., is Chief Scientist of Klein Associates, a group he formed in 1978 to better understand how to improve decision making in individuals and teams. Klein Associates (now a Division of Applied Research Associates, Inc.) has more than 35 employees working on projects for government (e.g., Army, Navy, Air Force, Marines) and commercial clients (e.g., Mead Johnson, Kodak, McKinsey, and Procter & Gamble). Dr. Klein is one of the founders of the field of Naturalistic Decision Making. His work on recognitional decision making has been influential for the design of new systems and interfaces, and for the development of decision training programs. He has extended his work on decision making to describe problem detection, option generation, sensemaking, planning and replanning. In order to perform research on decision making in field settings, Dr. Klein and his colleagues have developed new methods of Cognitive Task Analysis. Klein Associates has used Cognitive Task Analysis methods to study decision making in more than 60 domains, including firefighting, command and control, software troubleshooting, healthcare, and consumer purchasing. Dr. Klein has presented workshops on Cognitive Task Analysis and on Decision Skills Training to a wide variety of professionals in the U.S. and abroad.

Dr. Klein received his Ph.D. in experimental psychology from the University of Pittsburgh in 1969. He was an Assistant Professor of Psychology at Oakland University (1970-1974) and worked as a research psychologist for the U.S. Air Force (1974-1978). He has written more than 70 papers and co-edited three books. He is the author of Sources of Power: How People Make Decisions (1998, MIT Press), which has sold more than 20,000 copies and has been translated into four languages, and The Power of Intuition (2004, A Currency Book/Doubleday). His next book, “Working Minds: A practitioner’s guide to Cognitive Task Analysis” (Crandall, Klein and Hoffman) will be published by MIT Press in spring of 2006.

Holly C. Baxter, Ph.D., is a Senior Scientist at Klein Associates, a Division of Applied Research Associates, specializing in Instructional Design, Evaluation Metrics, Organizational Development, and Training. Her research interests include Knowledge Management and all phases of cognitively based training including cognitive needs analysis, design, development, implementation, and developing cognitive metrics for evaluating training initiatives ranging from classroom training to simulations. As a member of Klein Associates she has served as the project manager and technical lead on projects including developing effective training for enhancing situation awareness,
designing embedded training solutions for damage control personnel, developing evaluation metrics for simulation based training, identifying cognitive training requirements for Future Force environments, and developing and using knowledge management tools to capture tacit knowledge in the field and turn that knowledge into effective just-in-time training. Dr. Baxter holds a BA in Communication from the University of Dayton, a MA in Organizational Communication and Training from Indiana University, and a Ph.D. from Indiana University in Organizational Communication and Management with a focus on Training, Instructional Design, and Organizational Development.
Cognitive Transformation Theory:  
Contrasting Cognitive and Behavioral Learning

Gary Klein Ph.D., Holly C. Baxter, Ph.D.  
Fairborn, OH  45324-6362  
gary@decisionmaking.com, holly@decisionmaking.com

INTRODUCTION

How can you improve cognitive skills? The conventional mechanisms of practice, feedback, and accumulation of knowledge rarely apply to cognitive skills in the same way they apply to behavioral skills. In this paper we argue that cognitive learning requires a different concept of the learning process.

Traditional approaches to learning seem clear-cut. You identify what you want the student to learn. You provide the knowledge and present an opportunity to practice the skill or concept. You give feedback so the student can gauge whether the learning has succeeded. Educating students in behavioral skills appears to be simply a matter of practice and feedback.

This approach to learning relies on a storehouse metaphor. It assumes that the learner is missing some critical form of knowledge—factual information or procedures. The learner or the instructor defines what knowledge is missing. They add this knowledge via a course or a practice regimen or through simple study. They provide feedback to the learner. Then, they test whether the new knowledge was successfully added to the storehouse.

We believe that this storehouse metaphor does not work for cognitive skills. The storehouse metaphor may be useful for learning factual information or for learning simple procedures. But cognitive learning should help people discover new ways to understand events. We can distinguish different forms of knowledge that people need in order to gain expertise: declarative knowledge, routines and procedures, recognition of familiar patterns, perceptual discrimination skills, and mental models.

The storehouse metaphor seems best suited for acquiring declarative knowledge and for learning new routines/procedures. It may be less apt for building pattern-recognition skills. It is least appropriate for teaching people to make perceptual discriminations and for improving the quality of their mental models. When people build a larger repertoire of patterns and prototypes they are not simply adding new items to their list. They are learning how to categorize the new items and are changing categories and re-defining the patterns and prototypes as they gain new experience. The storehouse metaphor implies a simple additive process, which would lead to confusion rather than to growth. We encounter this kind of confusion when we set up a new filing system for an unfamiliar type of project and quickly realize that adding more files is only creating more confusion—the initial categories have to be changed.

When people develop perceptual discrimination skills they are learning to make distinctions that they previously didn’t notice. They are learning to “see the invisible” (G. A. Klein & Hoffman, 1993), in the sense that they can now make discriminations they previously didn’t notice. Perceptual learning depends on re-fashioning the way we attend and the way we see, rather than just adding additional facts to our knowledge base.

Cognitive skills depend heavily on mental models. We define a mental model as a cluster of causal beliefs about how things happen. According to Glaser and Chi (1988), mental models are used to organize knowledge. Mental models are also described as knowledge structures and schemata.

Cognitive learning is not simply a matter of adding additional beliefs. Rather, we have to revise our belief system as experience shows the inadequacy of our current ways of thinking. We discover ways to extend or even reject our existing beliefs in favor of more sophisticated beliefs.

How are mental models formed and modified? The storehouse metaphor of learning misses the sensemaking that is at the heart of learning cognitive skills. Students are not just acquiring new knowledge.
They are changing the way they see things and think about them. They are making sense of conflicting and confusing data.

The next section describes the kinds of sensemaking needed for cognitive learning. Following that, we present the concept of cognitive transformation as an alternative to the storehouse metaphor. Finally, we offer some implications for achieving cognitive learning.

SENSEMAKING REQUIREMENTS FOR LEARNING COGNITIVE SKILLS

What is hard about learning cognitive skills? None of the traditional components of learning—diagnosis, practice, feedback, or training objectives—are straightforward. Each of them depends heavily on sensemaking (e.g., Weick, 1995). Bloom’s (1956) taxonomy includes a component of synthesis—building a structure or pattern from diverse elements; and putting parts together to form a whole, with emphasis on creating a new meaning or structure. This corresponds to the process of sensemaking. We treat cognitive learning as a sensemaking activity that includes four components: diagnosis, learning objectives, practice, and feedback.

a) Diagnosis. Diagnosing the reasons for weak performance depends on sensemaking. The instructor has to ferret out the reasons why the student is confused and making errors. Sometimes students don’t even notice errors or weaknesses, and may resist suggestions to overcome problems they don’t realize they have. Even if students do realize something is wrong, the cause/effect mechanisms are subtle and complex. Outcome feedback usually does not provide any clues about what to do differently. That is why instructors need to provide immediate process feedback as the student progresses through the learning process, but they first must diagnose what is wrong with the students’ thinking. Diagnosing the reason for poor performance is a challenge to students. It is also a challenge to the instructors who may not be able to figure out the nature of the problem.

Diagnosis is a challenge to instructional developers. The classical systems approach to instructional design is to subtract the existing Knowledge, Skills, and Abilities (KSAs) from the needed KSAs. But for cognitive skills, instructional developers need to understand why the students are struggling. The goal of diagnosis goes beyond establishing learning objectives—it depends on discovering what flaw in a mental model needs to be corrected.

b) Learning objectives. With the storehouse metaphor, learning objectives are clear and succinct—the additional declarative or procedural knowledge to be imparted, and the changes in performance that reflect whether the student has acquired the new material.

c) But for cognitive learning, the objectives may be to help the students revise their mental models and perhaps to re-organize the way they categorize events. Some learning theorists suggest that it is more about how new learning is integrated with what’s already known that’s most important, along with the affordances this new learning represents. For example, both Kolb (1984) and Dewey (1938) focus on learning through experience. What’s important in Kolb’s reflective observation stage is how the learner transforms an experience into learning through reflection. During reflection, the student compares the new learning to what’s already known and tries to make it fit with existing knowledge and see how to leverage this new knowledge for additional learning.

For Dewey, the key is what the learner does with experience. Not all experiences are equal, and not all experiences are educational. According to Dewey, individuals reflect on their experiences to learn what thoughts and actions can change real-world conditions that need improving. You leverage your existing knowledge, your new learning (through reflection or “thinking”), along with the knowledge of others, to identify opportunities to improve the current, problematic situation. Dewey thought that people were constantly trying to resolve perplexing intellectual situations and difficult moral situations.

For theorists such as Kolb and Dewey, accumulating or storing knowledge is not an end-state. Instead, it kicks off a series of cognitive activities by the individual to figure out ways to test
the “goodness” of the new learning through active experimentation or to use the new learning to change an unsatisfactory situation.

We assert that novices may not have mental models for an unfamiliar domain and will struggle to formulate even rudimentary mental models linking causes to effects. Their learning objective is to employ sensemaking to generate initial mental models of cause/effect stories, whereas experts are revising and adding to current mental models.

d) Practice. Providing students with practice is necessary for gaining proficiency. But with cognitive skills, practice is not sufficient. For cognitive skills, students often may not know what they should be watching and monitoring. They need adequate mental models to direct their attention, but until they get smarter, they may fail to spot the cues that will help them develop better mental models.

Managing attention depends on sensemaking. Feedback won’t be useful if the trainee doesn’t notice or understand it—and that requires the trainee to know what to attend to and when to shift attention. Barrett, Tugade and Engle (2004) have suggested that attention management accounts for much of the individual differences in working memory—the ability to focus attention and not be distracted by irrelevancies. For these reasons, we argue that effective practice depends on attention management: seeking information—knowing what to seek and when to seek it—and filtering distracting data.

e) Feedback. Providing students with feedback will not be useful if they do not understand it. For complex cognitive skills, such as leadership, time lags between actions and consequences will create difficulties in sorting out what worked, what did not work, and why. Learners need to engage in sensemaking to discover cause-effect relationships between actions taken at time one and the effects seen at time two. Adding to the complication, learners often have to account for other actions and events that are interspersed between their actions and the consequences. They have to figure out what really caused the consequences, versus the coincidental events that had nothing to do with their actions. They have to understand the causes versus the symptoms of deeper causes, and they have to sort out what just happened, the factors in play, the influence of these factors, and time lags for effects.

To add to these complications, having an instructor provide feedback can actually get in the way of transfer of learning (Schmidt & Wulf, 1997), even though it increases the learning curve during acquisition. By placing students in an environment where they are given rapid feedback, the students are not compelled to develop skills for seeking their own feedback. Further, students may become distracted from intrinsic feedback because it is so much easier to rely on the extrinsic feedback. As a result, when they complete what they set out to learn, they are not prepared to seek and interpret their own feedback.

For cognitive learning, one of the complications facing instructors is that the flawed mental models of the students act as a barrier to learning. Students need to have better mental models in order to understand the feedback that would invalidate their existing mental models. Without a good mental model, students will have trouble making use of feedback, but without useful feedback, students will not be able to develop good mental models. That is why cognitive learning may depend on unlearning as well as learning.

**The Process of Unlearning**

For people to develop better mental models they may have to unlearn some of their existing mental models. The reason is that as people gain experience, their understanding of a domain should become more complex and nuanced. The mental models that provided a rough approximation need to be replaced by more sophisticated ones. But people may be reluctant to abandon inadequate mental models, as they may not appreciate the inadequacies. They may attempt to explain away the inconsistencies and anomalies. A number of researchers have described the reluctance to discard outmoded mental models even in the face of contrary evidence. DeKeyser and Woods (1993) have commented on the way decision makers fixate on erroneous beliefs. Feltovich, Spiro, and Coulson (1997) used a garden path paradigm and identified a range of knowledge shields that pediatric cardiologists employed to discount inconvenient data. Chinn and Brewer (1993) showed that scientists and science students alike deflect inconvenient data. Klein, Phillips, Rall, and Peluso, (2006) described the “spreading corruption” that resulted when people distorted data in order to retain flawed mental models.
As people become more experienced, their mental models become more sophisticated and therefore become more effective in explaining away inconsistencies. Fixations should become less tractable as cognitive skills improve. Therefore, people may have to unlearn their flawed mental models before they can acquire better ones. This is one of the main reasons that contrary to common belief, it actually more difficult for experts to learn new material than it is for novices. Sensemaking here is a deliberate activity to discover what is wrong with one’s mental models and to abandon and replace them.

The process of unlearning has been studied more at the organizational level, than at the individual level. In some ways, the process of unlearning that we are presenting resembles the scientific paradigm replacements described by Polanyi (1958) and Kuhn (1996). The term “unlearning” is widely used in the field of organizational learning. Starbuck and Hedberg (2001) stated that “Organizations’ resistance to dramatic reorientations creates a need for explicit unlearning...Before attempting radical changes, [organizations] must dismantle parts of their current ideological and political structures. Before they will contemplate dramatically different procedures, policies, and strategies, they must lose confidence in their current procedures, policies, strategies, and top managers.” (p. 339). We believe that these observations apply to individuals as well as to organizations, and that the concept of unlearning needs to become part of a cognitive learning regimen.

Individuals also resist changing their mental models. Although we are not aware of explicit research on unlearning at the individual level, we have found examples where this practice was put into use. DiBello and her colleagues developed a two-day training program to help managers think more effectively about their work (for a fuller description, see G. Klein, 2004). The first day was spent in a simulation of their business designed to have the managers fail in the same ways they were failing in real-life. This experience helped the managers lose confidence in their current mental models of how to conduct their work. The second day gave the managers a second shot at the simulated exercise and a chance to develop and use new mental models of their work. Schmitt (1996) designed similar experiences for the U.S. Marine Corps. His Tactical Decision Games—low fidelity paper and pencil exercises—put individual Marines into situations that challenged their thinking and made them lose confidence in their mental models of tactics and leadership. The exercises provided a safe environment for re-thinking some of their closely-held beliefs.

A number of researchers, e.g., Rouse and Morris (1986) have voiced concerns about invoking the notion of mental models. The concept of a mental model is typically so vague and ambiguous that it has little theoretical or applied value. However, Klein and Hoffman (in preparation) have countered this criticism by presenting a more detailed description of mental models. They argue that the term “mental model” is an umbrella that covers a variety of relationships: causal, spatial, organizational, temporal, and so forth. As long as we are clear about which type of relationship we are interested in, much of the murkiness of “mental models” disappears.

With regard to cognitive learning, our emphasis is usually on causal relationships. During the learning process, people are engaged in sensemaking to understand and explain how to make things happen. Under the right circumstances, they may also discover better ways to think about causal connections.

People have to diagnose their performance problems, manage their attention, appreciate the implications of feedback, and formulate better mental models by unlearning inadequate models. Learners are not simply accumulating more knowledge into a storehouse. They are changing their perspectives on the world. That is why we hypothesize that these changes are uneven, rather than smooth and cumulative. The next section presents a theoretical account of the transition to high levels of expertise.

**COGNITIVE TRANSFORMATION THEORY**

In this section we present an account of the transition process for acquiring cognitive skills. We are primarily interested in how people learn better mental models to achieve a stronger understanding of what has been happening and what to do about it. In contrast to a storehouse metaphor of adding more and more knowledge, we offer the notion of cognitive transformation—that progress in cognitive skills depends on successively shedding outdated sets of beliefs and adopting new beliefs, the way a snake sheds its skin. We call this account of cognitive learning “Cognitive Transformation Theory” (CTT).

Our central claim is that conceptual learning is discontinuous, rather than smooth. We make periodic advances when we replace flawed mental models with
better ones. However, during the process of cognitive development our mental models get harder to disconfirm. As we move further up the learning curve or have more expertise, we have to put more and more energy into unlearning—disconfirming mental models—in order to accept better ones.

We do not smoothly acquire knowledge, as in a storehouse metaphor. Our comprehension proceeds by qualitative jumps. However, at each juncture our new mental models direct what we attend to and explain away anomalies. As a result, we have trouble diagnosing the flaws in our thinking. We have trouble taking advantage of feedback. Because of problematic mental models, people often misdiagnose their limitations and discard or misinterpret informative feedback. The previous mental model, by distorting cues and feedback, acts as a barrier to advancement. So progress is not simply in fits and starts. Progress may involve some backtracking to shed mistaken notions. In addition, flawed beliefs have also influenced the way people encoded experiences in the past. Simply changing one’s beliefs will not automatically change the network of implications generated from those beliefs. As a result, people may struggle with inconsistencies based on different mental models that have been used at different times in the past.

In order to move forward, we need to unlearn our old mental models in order to adopt new and better ones. Once people replace an inadequate mental model with a better one, then they can see all kinds of patterns. Instructional developers may have to take responsibility for designing interventions that help students unlearn their flawed mental models.

We can represent cognitive transformation theory as a set of postulates:

- Mental models are central to cognitive learning: instruction needs to diagnose limitations in mental models, design interventions to help students appreciate the flaws in their mental models, and provide experiences to enable trainees to discover more useful and accurate mental models.
- Mental models are modular. People have a variety of fragmentary mental models, and they weave these together to account for a novel observation. In many cases people rely on just-in-time (JIT) mental models. People are usually not matching events to sophisticated theories they have in memory. They are using fragments and partial beliefs to construct relevant mental models. For most domains, the central mental models describe causal relationships. They describe how events transform into later events. They describe ways of achieving desired transformations. Causal mental models typically take the form of a story.
- Experts have more sophisticated mental models in their domain of practice than novices. Experts have more of the fragmentary beliefs needed to construct a plausible JIT mental model. Experts also have more elaborate mental models to begin with. Therefore, they are starting their construction from a more advanced position. Finally, experts have more accurate causal mental models and have tested and abandoned more inadequate beliefs.
- Experts build their repertoires of fragmentary mental models in a discontinuous fashion. In using their mental models, experts distort data, oversimplify, explain away diagnostic information, and misunderstand events. At some point, experts realize the inadequacies of their mental models. They abandon their existing mental models and replace these with a better set of causal beliefs. And the cycle begins again.
- Learning curves are usually smooth because researchers combine data from several subjects. The reason for the smoothness is the averaging of discontinuous curves. At any given point, learners are either elaborating their current beliefs or are preparing to shift to a new and better set of causal beliefs.
- Experts are fallible. No set of mental models is entirely accurate and complete.
- Knowledge shields are the set of arguments learners can use to explain away data that challenges their mental models (Feltovich et al., 1997). Knowledge shields pose a barrier to developing cognitive skills. People are skilled at holding onto cherished beliefs. The greater the expertise, the easier it is to find flaws in disconfirming evidence. The better the mental models, the easier it is to use them to discount anomalous observations. The S-shaped learning curve reflects the increasing difficulty of replacing mental models as people’s mental models become more accurate.
- Knowledge shields affect diagnosis. Active learners try to overcome their limitations. But they need to understand what those limitations are. Knowledge shields, based on poor mental models, can lead learners to the wrong diagnoses of their poor performance.
- Knowledge shields affect feedback. In building mental models about complex situations, people
receive a lot of feedback. However, the knowledge shields enable people to discard or neutralize contradictory data.

- Progress depends on unlearning. The better the causal models, the more difficult it is to discover their weaknesses and replace them. In many cases, learners have to encounter a baffling event, an unmistakable anomaly, or an intelligent failure in order to begin doubting their mental models. They have to lose faith in their existing mental models before they can review the pattern of evidence and formulate a better mental model. People can improve their mental models by continually elaborating them, by replacing them with better ones, and/or by unlearning their current mental models. Cognitive development relies on all three processes.

Cognitive Transformation Theory generates several testable hypotheses. It asserts that individual learning curves will be discontinuous, as opposed to the smooth curves found when researchers synthesize data across several subjects. Cognitive Transformation Theory suggests a form of state-dependent learning. The material learned with one set of mental models may be inconsistent with material learned with a different mental model. Consequently, learners may be plagued with inconsistencies that reflect their differing beliefs during the learning cycle. Cognitive Transformation Theory suggests that teaching new cognitive skills to low-intermediate trainees may sometimes be easier than teaching them to high-intermediate trainees. The less-skilled trainees may not have to unlearn flawed mental models.

CONCLUSIONS

Now we can see what is wrong with the storehouse metaphor of learning described at the beginning of this paper. Learning is more than adding additional information. Learning is about changing the way we understand events, changing the way we see the world, changing what counts as information in the first place. The functions of diagnosis, practice, and feedback are all complex and depend on sensemaking.

To replace the storehouse metaphor we have presented a theory of cognitive transformation. We claim that cognitive skills do not develop as a continual accumulation. Rather, cognitive skills and the mental models underlying them progress unevenly. Flawed mental models are replaced by better ones but the stronger the mental models the more difficult to dislodge them. As a result, learners explain away anomalies, inconsistencies, inconvenient feedback, and misdiagnose their problems. How we teach cognitive skills, therefore, has to help people unlearn their current mental models before helping them develop better ones. If this unlearning process doesn’t occur, the students will use their current mental models to discount the lessons and the feedback.

Cognitive Transformation Theory may offer a shift in perspective on cognitive learning. It relies on sensemaking as the core function in learning cognitive skills, as opposed to a storehouse metaphor. Therefore, researchers should be able to formulate different hypotheses about which types of learning interventions will be effective and what dependent variables to measure.

ACKNOWLEDGEMENTS

The authors would like to thank Joseph Cohen for his support of this project developed under Contract M67854-04-C-8035 (issued by MARCORSYSCOM/PMTRASYS). We would also like to thank Sterling Wiggins, Dr. Karol Ross, and Jennifer Phillips for their valuable critiques and inputs.

REFERENCES


