PTG GLOBAL
ABN 27 089 738 205
L16, 207 Kent St
Sydney NSW 2000

P 61 2 9251 4200
1300 660 973
F 61 2 9251 7422

ptg-global.com

1. Cognitive load theory and user interface design: Making software easy to learn and use (Part 1)

1.1. Executive summary

Cognitive load theory (CLT) is an instructional design theory with the aim of assisting instructional designers to reduce the load caused by poor design of the learning materials.

CLT uses an information processing approach to cognition, involving working memory and long term memory. Long term memory stores knowledge and skills, while working memory performs the intellectual activities associated with thinking and processing information. Knowledge and skills are transferred to long term memory only after they have been attended to and processed by working memory. However, working memory is limited in its capacity and duration (Cooper, 1998).

Therefore, when the design of the instructional material itself causes cognitive demands on the user, working memory capacity is reduced which makes learning the material more difficult. The same applies to the design of software. When the software causes undue demands on the user, as a result of:

- Being poorly designed such that it is unclear how to use the application,
- The data or settings to be remembered are spread across multiple screens, or;
- The way the software works conflicts with existing knowledge about the domain

These artefacts will make it more difficult for the user to understand the material, making learning difficult or even impossible.

This paper provides an overview to cognitive load theory and its role in guiding user interface design. Its purpose is to:

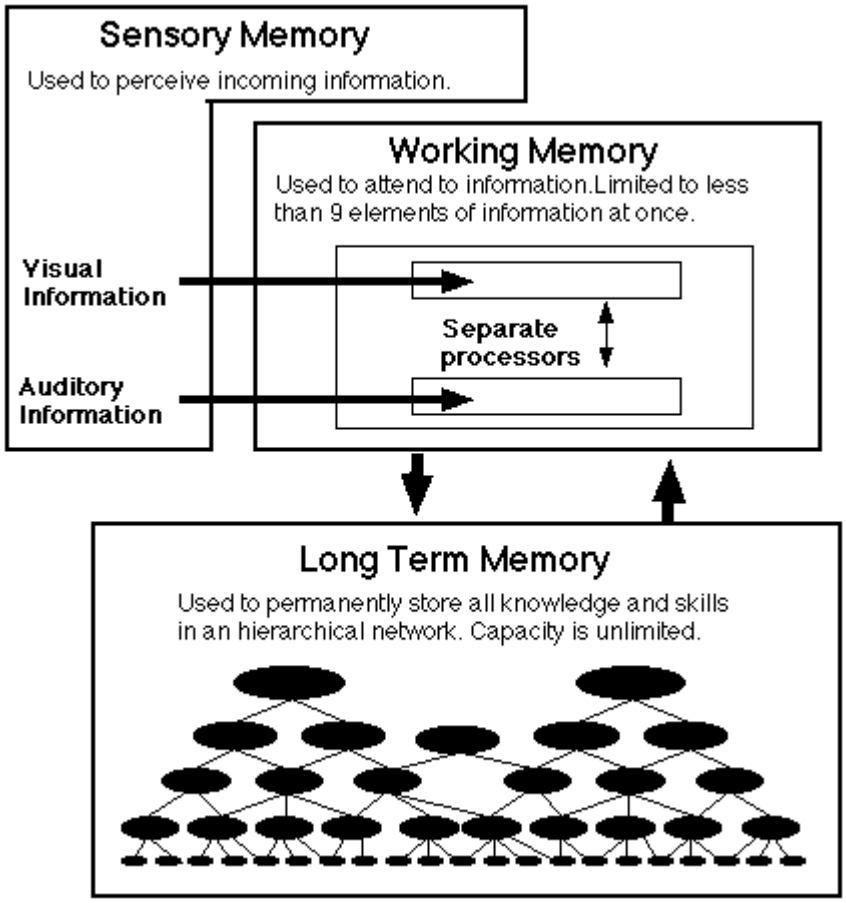
- Provide the reader with an understanding of how people process information,
- The artefacts that cause cognitive load, and
- The general principles behind reducing cognitive load so there are more working memory resources available for learning and activities.

Part 2 of this paper will focus on how to use CLT to:

- Make software easy to learn,
- Present content in the right way to support information processing by the user,
- Provide just in time training where the UI design teaches the user how to use the software and perform their job.

1.2. The nature of Memory

The modal model memory (Cooper, 1998) distinguishes between three memory types (modes). These are the sensory memory, the (short-term) working memory and the long-term memory. These modes define an information-processing model of human cognitive architecture in an integrated way. The following figure illustrates this:



(From Cooper, 1998, http://education.arts.unsw.edu.au/CLT_NET_Aug_97.HTML).

1.2.1. Sensory memory

The sensory memory deals with stimuli that are processed through our senses. These can be sights, sounds, smells, touches, and tastes. These memories extinguish quickly – about half a second for visual information and 3 seconds for auditory information. Unless the sensory information is attended to, i.e. identified, classified and assigned meaning to, it will be forgotten. The content of the sensory memory is constantly overwritten by new input. When you close your eyes, the visual image vanishes quickly, although any afterimage is your visual short-term memory. This overwriting mechanism is necessary because of the vast quantity of data in an image and the continuous changes in images. This has implications for graphical user interfaces and multimedia: if images are not held on the screen long enough, we will not be able to extract much information from them.

The content of the sensory memory is still abstract with no meaning attached to its input. Meaning is generated when the input reaches the central cognitive short-term memory for interpretation. The cognitive processor is responsible for object identification. The cognitive processor has an associated short-term memory used for storage of temporary working information. This information can be extracted from the sensory processors or the long-term memory. In the modal memory model, all the short-term memories are referred to as working memory. The cognitive processor performs most of the 'thinking' activity. The results of thinking can either be placed back in short-term memory, stored in long-term memory or be passed on to the motor processor to elicit behaviour.

1.2.2. Working memory

The term "working" memory implies that it is more than a passive, temporary storing place for leftover perception – as once was believed to be more or less the case for the old concept of short-term memory. The contents of working memory can be combined with stored knowledge from the long-term memory and manipulated, interpreted and recombined to develop new knowledge, form goals, and assist learning and interaction with the physical world (Logie, 1999).

The working memory or the short-term working memory is equivalent to computer RAM, i.e. the working memory of the central processor. In contrast to computers, the human working memory has a low capacity; it loses its content unless being refreshed every 200ms. The read/write access time is quite quick (about 70 ms) which means that information can be held in working memory by continual rewriting.

The working memory can typically hold 7±2 items for rehearsal (Miller, 1956, in Sutcliffe, 1995). It will rapidly decay if we don't do anything special to it to keep it active. Instead of storing information in 'bytes', as in computers, it is stored in chunks of information. For example, it is common practice to combine phone numbers into chunks rather than listing all digits in one sequence. Consider remembering the phone number 9237 9154 as opposed to 9 2 3 7 9 1 5 4. The former number may be easier to remember than the latter.

The chunks of information can vary from simple characters and numerals to more complex abstracts and images. The working memory can be expanded by abstracting qualities from the basic information and store the abstraction instead. Chunking need not be based upon any underlying meaning or logic within the elements of the material. However, if there is an underlying meaning/logic that can be identified, the learning and memory is greatly enhanced (Cooper, 1998). In general, the more order that can be imposed on the raw data the better the chunking.

1.2.3. Long term memory

Long-term memory (LTM) refers to the immense amount of knowledge and skills that we hold in a more or less permanently accessible form. Retrieval of facts from the LTM can be remarkably fast, especially for frequently used items. For example, it doesn't take long to recall our name, date of birth, letters in the alphabet and so on. For less frequently used information, retrieval time can be longer (Cooper, 1998).

Retrieval can be a quite complex process. Often, remembering occurs minutes after you made the original effort to retrieve it. During this intervention, attention would have been devoted to other matters; hence it appears that a background memory processor is invoked to effect difficult memory searches. According to the information-processing model, the retrieval process is simply a function of the cognitive processor. Memory seems to be activated by use, so frequently or recently used items are easier to recall. In these situations, both recognition and recall happen quickly and instantly. However, in the "tip of the tongue" situations, there is a noticed difference between the activation of the memory trace by cues (recognition) and the actual retrieval of the information (recall). In recognition and recall there is also evidence of a "spreading activation" as remembering one fact often helps the recall of other related items (Sutcliffe,1995).

The huge capacity of LTM to store associations between complex configurations and consequent actions, and to store complex associative networks, such as categorisation skills and sequential procedures has implications for instructional design. For example, it should facilitate acquisition of domain specific KSA (knowledge, skills and abilities) rather than highly general reasoning skills.

1.3. Cognitive load theory, working memory and learning

1.3.1. Defining cognitive load

Cognitive load refers to the total amount of mental activity that the working memory has to attend to at an instance in time. The focus is on the role of working memory in the learning process. The number of elements that is imposed on working memory is the major contributor to cognitive load.

For example:

the statement	9 2	has a cognitive load of 2
the statement	7 9 5 3	has a cognitive load of 4
the statement	3 9 2 4 6 7 1 5	has a cognitive load of 8
the statement	3 9 3 5 7 1 5 0 3 5 1 8 6 2 4 1	has a cognitive load of 16

The measure used for cognitive load does not equate mathematical task difficulty; despite that statement 2 has doubled the cognitive load from statement 1, it is almost as easy to remember. The same principal yields for statements 3 and 4; statement 4 appears to be more than twice as difficult to remember as statement 3.

The most famous statement of cognitive load is Miller's (1956) research into the capacity of 7± 2 items. That is, our memory capacity, usually tested by a series of digits is around 7 items.

This principle is also mis-used in navigation design, with some people interpreting it to mean that website navigation can not have any more than 7 items in it. This is simply not true, because there is no demand on the user to actual remember the navigation. We can easily scan it again to refresh our memory of what was there. Even people who

use speech readers can cope with a larger number of navigation items, because they compensate for visual issues with a better working memory.

The principle does have relevance, however, with auditory interfaces (such as telephone IVR systems), where the caller must remember the spoken menu items and process them to determine their relevance to the task at hand. In such situations, the number of items should be limited to 5, allowing some resources to process the items.

1.3.2. Working Memory

Working memory is important because it is here that we organise and process information when we are learning. Information is 'sorted' in working memory and organised into a relevant schema. Schemas can be understood as models or hypothetical structures that organise our knowledge of the world. Experts in fields such as biochemistry or mathematics have more extensive schemas for their area of interest, and can therefore better organise task-relevant information than novices. For example, the digit span 3.14285714 would be extremely difficult for a novice to remember if they were shown it for 5 seconds. A mathematics expert, however, would access their 'maths schema' and immediately recognise the span as the first 9 digits of pi, increasing the likelihood that they would remember the sequence. Since novices do not have access to such an elaborate schema, they must attempt to hold the entire span in working memory.

1.3.3. Learning

Learning is the process by which information (in terms of knowledge and skills) is encoded into long term memory, so they can be retrieved and applied at a later date (Cooper, 1998). Encoding takes place in working memory where relationships are created and content rehearsed.

The information in our long term memory is stored in schemas. Practice, experience and formal education contribute to the differences in people's schemas, especially about the same knowledge areas. People who specialise in an area have a deeper, richer and more complex schema than do people who are only aware of the basics.

Learning is most successful when the new information is clearly related to existing schemas. If we already know something about a topic, then learning new information is easier. If the new information conflicts with what we know, then learning can be harder. If information is presented within structures than are unrelated to the domain, or even without structure, then learning is also harder. For example, telling someone to perform a sequence of actions in a certain way, without evidence of rhyme or reason, will be very difficult and we would have to rote learn the material.

1.3.4. Types of cognitive load for working memory

Intrinsic cognitive load

Intrinsic cognitive load is determined by the intrinsic nature (difficulty) of the to-be-learned content. This can not be modified or improved by instructional design. Regardless of how it is presented it retains its inherent level of element interactivity.

Extraneous cognitive *load*

Extraneous cognitive load is due to the instructional materials used in the presentation of the information. As opposed to intrinsic load, this can be modified and manipulated by the instructional design to facilitate learning. An example of high extraneous load can be when the leaner must extract information from multiple sources and integrate it, such as glossary of technical terms separated from the text.

When the intrinsic load is low the mental resources are less "burdened" and more working memory should be available to learn from nearly any type of instructional material, even if the extraneous load is high. For example, if the to-be-learned material has simple content (i.e. low intrinsic load), it is likely to be learned and understood even if the way it is presented is difficult (i.e. high extraneous load). However, if both intrinsic and extraneous cognitive loads are high, the total cognitive load will exceed the mental resources, which may result in failure to learn. Conversely, if the extraneous load is low the chances of learning difficult material (which has high intrinsic load) will increase.

1.4. The difference between novices and experts

For any given cognitive domain, the differences between being a novice and an expert come down to differences in schema expansiveness and level of automation. These two components (schemas and automation) appear to explain all expert-novice differences. Having an expansive set of schemas means that the person (the expert) has seen almost every possible situation in the content domain earlier. As such, they have learnt what response is required for each situation, and the responses can be carried out automatically. Experts are merely going through routine exercises.

The greater our expertise is in one particular domain, the greater is the working memory capacity for information in that domain. Thus, chess players can retain details of chess games played simultaneously even when blindfold (Saariluoma, 1995) and soccer supporters can remember scores from matches more accurately than more casual fans (Morris, Tweedy, and Gruneberg, 1985). In each case, having expert knowledge allows efficient coding and retrieval of information within the area. These memory skills clearly rely on the working memory, but expertise greatly facilitates activation of relevant information in the knowledge base, and this activated knowledge offers significant support for the limited working memory.

Novices, on the other hand, have relatively few schemas. This makes it difficult to recognise anything but basic and common situations where they can draw upon earlier experiences. They must solve a problem every time they encounter the content domain. In addition, when they realise what response is the correct one, they may have problems performing the response. In contrast to the above group, the way the learning material is presented for novices (the instructional design) can have significant implications for the actual learning outcome (Cooper, 1998).

The differences in expertise in the domain area can have usability implications. Website audiences may vary from one extreme to the other in expertise and it can reach a much bigger audience than those who are the typical users of the product. The designer should therefore not assume that the site user is the same as the typical user who knows the product well. For example, navigation can became a problem because the site seemed to assume that the user had specific domain knowledge – or knowledge about the specific business area. However, when the users are unfamiliar with the domain, they don't understand the options that are presented to them. your users can have difficulty if they can't distinguish between similar sounding concepts such as Fidelity

Mutual Funds or Fidelity Dailt NAVs. Here, designers have assumed that users would have a schema for the finance domain, and the site's own products, when they did not. The site navigation therefore confused them and made browse tasks with the site very difficult.

1.4.1. Why things can be hard to learn

CLT tells us that some material is easier to learn than others. This occurs when:

- An item can be understood in isolation, then it is easy to learn. For example, learning a foreign word is quite easy, because you can simply learn them and they don't depend on others. Learning grammar is harder because all the words in a sentence must agree to make sense.
- The number of items to be learned exceeds the capacity of working memory (i.e. 7±2 items).
- When there are no resources left for processing the information in working memory to draw meaning from them. For example, if we are to learn the people's names, then the best way is to make up a story about them. If we don't have time or don't do it, or just try to quickly rote learn a large number of names, then learning will generally be poor.
- When information is provided in different formats (e.g. text and graphics), and the task of integrating the two takes resources away from actually learning the principles. This is especially the case when text and graphics present information that has to be integrated before the concept can be properly understood. The process of integration takes up working memory load, which makes the un-integrated material harder to learn. This is referred to in CLT literature as the 'split attention affect'.
- Using redundant material, such as text and graphics, where the text or graphics are sufficient to stand alone. Learning is reduced because working memory capacity is taken up trying to integrate the information, even though both sources provide the same information (and so the integration is unnecessary and second source redundant). A single source of instruction yields superior performance.

1.5. Conclusion

In summary, many of the CLT principles are counter-intuitive, especially the observation that referring to a manual while learning software is a sub-optimal learning strategy. Interestingly, when learning software, students referring to the manual alone (assuming it is has been developed using CLT principles), and then using the software afterwards, outperform students working through the manual while using the software (Cerpa, Chandler & Sweller, 1996).

CLT's importance for usability cannot be understated; tapping in to the ways that people naturally learn allows them to free up their mind to focus on the message that you're trying to get across, rather than wasting time and mental energy integrating information that has been badly presented. Designers would do well to incorporate principles of cognitive psychology – and cognitive load theory in particular – to make the user's experience efficient and user friendly.

CLT shows that conventional wisdom regarding instructional methods is heavily flawed. CLT can be experimentally shown, using real-world situations, to have significant benefits over conventional instructional techniques in terms of:

- Reducing training time,
- Improved performance on tasks using the learned knowledge / skills,
- Improved performance on other tasks where the learned information can be applied (i.e. improved transfer).

Part 2 of this article will focus on how to use CLT to:

- Make software easy to learn
- Present content in the right way to support information processing by the user,
- Provide just in time training where the UI design teaches the user how to use the software and perform their job.

1.6. References

Ayres, P. (1993). Why goal free problems facilitate learning. *Contemporary Educational Psychology*, 18, 376-381

Cerpa, N., Chandler, P., & Sweller, J. (1996). Some conditions under which integrated computer-based training software can facilitate learning. *Journal of Educational Computing Research, 15,* 345-367

Cooper, G. (1998). Research into Cognitive load Theory and Instructional Design at UNSW. Dr. Graham Cooper, School of Education Studies, The University of New South Wales. http://education.arts.unsw.edu.au/CLT_NET_Aug_97.HTML

Della Sala, S, and Logie, R.H. (1993) When working memory does not work: The role of working memory in neuropsychology. In Logie, R.H. (1999). Working memory. *The Psychologist, 12(4),* 174-178

Ellis, N.C. & Hennelley, R.A. (1980). A bilingual word length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology, 71,* 43-52

Morris, P.E., Tweedy, M. and Gruneberg, M.M. (1995). Interest, knowledge, and the memorization of soccer scores. *British Journal of Psychology, 76,* 415-425

Saaarilluoma, P. (1995). Chess Players' Thinking. London: Routledge in Logie, R.H. (1999). Working memory. *The Psychologist, 12,(4)*, 174-178

Stigler, J.W., Lee, S.Y. and Stevenson, H.W. (1986). Digit memory in Chinese and English: Evidence for a temporally limited store. *Cognition*, 24, 1-20.

Sutcliffe, A.G. (1995). *Human computer interface design*. London, McMullan.

2. About the primary Author

Craig is the founder and Managing Director of The Performance Technologies Group (PTG Global), with over 15 years in user experience, user interface design and change management.

Craig runs the R&D function at PTG, having produced a number of world firsts including XPDesign – the first systematic methodology for user interface design and Certified Usable – the first guarantee for usability and user experience.

Craig has been the primary architect behind many of Australia's most popular websites including CBA, Virgin Blue and ASIC and works on cutting edge technologies such as touch, medical and special-purpose applications.

Craig holds a Masters qualification in organisational psychology, is a member of the APS and the APS College of Organisational Psychologists and is a Registered Psychologist in NSW. He is also an Associate of the University of NSW and Macquarie University.

Contact Craig on:

Email: craige@ptg-global.com

Phone: +61 (0)2 9251 4200

Mobile: +61 (0) 416 266 216

Address: Level 16, 207 Kent St, Sydney, NSW, 2000, Australia