Chapter 3

Understanding users

3.1 Introduction

Imagine trying to drive a car by using just a computer keyboard. The four arrow keys are used for steering, the space bar for braking, and the return key for accelerating. To indicate left you need to press the F1 key and to indicate right the F2 key. To sound your horn you need to press the F3 key. To switch the headlights on you need to use the F4 key and, to switch the windscreen wipers on, the F5 key. Now imagine as you are driving along a road a ball is suddenly kicked in front of you. What would you do? Bash the arrow keys and the space bar madly while pressing the F4 key? How would you rate your chances of missing the ball?

Most of us would balk at the very idea of driving a car this way. Many early video games, however, were designed along these lines: the user had to press an arbitrary combination of function keys to drive or navigate through the game. There was little, if any, consideration of the user's capabilities. While some users regarded mastering an arbitrary set of keyboard controls as a challenge, many users found them very limiting, frustrating, and difficult to use. More recently, computer consoles have been designed with the user's capabilities and the demands of the activity in mind. Much better ways of controlling and interacting, such as through using joysticks and steering wheels, are provided that map much better onto the physical and cognitive aspects of driving and navigating.

In this chapter we examine some of the core cognitive aspects of interaction design. Specifically, we consider what humans are good and bad at and show how this knowledge can be used to ***inform*** the design of technologies that both ***extend*** human capabilities and ***compensate*** for their weaknesses. We also look at some of the influential cognitively based conceptual frameworks that have been developed for explaining the way humans interact with computers. (Other ways of conceptualizing human behavior that focus on the social and affective aspects of interaction design are presented in the following two chapters.)

The main aims of this chapter are to:

- Explain what cognition is and why it is important for interaction design.
- Describe the main ways cognition has been applied to interaction design.
- Provide a number of examples in which cognitive research has led to the design of more effective interactive products.
- Explain what mental models are.
- Give examples of conceptual frameworks that are useful for interaction design.
- Enable you to try to elicit a mental model and be able to understand what it means.

3.2 What is cognition?

Cognition is what goes on in our heads when we carry out our everyday activities. It involves cognitive processes, like thinking, remembering, learning, daydreaming, decision making, seeing, reading, writing and talking. As Figure 3.1 indicates, there are many different kinds of cognition. Norman (1993) distinguishes between two general modes: experiential and reflective cognition. The former is a state of mind in which we perceive, act, and react to events around us effectively and effortlessly. It requires reaching a certain level of expertise and engagement. Examples include driving a car, reading a book, having a conversation, and playing a video game. In contrast, reflective cognition involves thinking, comparing, and decision-making. This kind of cognition is what leads to new ideas and creativity. Examples include designing, learning, and writing a book. Norman points out that both modes are essential for everyday life but that each requires different kinds of technological support.

Figure 3.1 What goes on in the mind?

Cognition has also been described in terms of specific kinds of processes. These include:

- attention
- perception and recognition
- memory
- learning
- reading, speaking, and listening
- problem solving, planning, reasoning, decision making

It is important to note that many of these cognitive processes are interdependent: several may be involved for a given activity. For example, when you try to learn material for an exam, you need to attend to the material, perceive, and recognize it, read it, think about it, and try to remember it. Thus, cognition typically involves a range of processes. It is rare for one to occur in isolation. Below we describe the various kinds in more detail, followed by a summary box highlighting core design implications for each. Most relevant (and most thoroughly researched) for interaction design is memory, which we describe in greatest detail.

Attention is the process of selecting things to concentrate on, at a point in time, from the range of possibilities available. Attention involves our auditory and/or visual senses. An example of auditory attention is waiting in the dentist's waiting room for our name to be called out to know when it is our time to go in. An example of attention involving the visual senses is scanning the football results in a newspaper to attend to information about how our team has done. Attention allows us to focus on information that is relevant to what we are doing. The extent to which this process is easy or difficult depends on (i) whether we have clear goals and (ii) whether the information we need is salient in the environment:

(i) ***Our goals*** If we know exactly what we want to find out, we try to match this with the information that is available. For example, if we have just landed at an airport after a long flight and want to find out who had won the World Cup, we might scan the headlines at the newspaper stand, check the web, call a friend, or ask someone in the street.

When we are not sure exactly what we are looking for we may browse through information, allowing it to guide our attention to interesting or salient items. For example, when we go to a restaurant we may have the general goal of eating a meal but only a vague idea of what we want to eat. We peruse the menu to find things that whet our appetite, letting our attention be drawn to the imaginative descriptions of various dishes. After scanning through the possibilities and imagining what each dish might be like (plus taking into account other factors, such as cost, who we are with, what the specials are, what the waiter recommends, whether we want a two- or three-course meal, and so on), we may then make a decision.

(ii) ***Information presentation*** The way information is displayed can also greatly influence how easy or difficult it is to attend to appropriate pieces of information. Look at Figure 3.2 and try the activity. Here, the information-searching tasks are very precise, requiring specific answers. The information density is identical in both

South Carolina

City	Motel/Hotel	Area code	Phone	Rates Single	Rates Double
Charleston	Best Western	803	747-0961	$26	$30
Charleston	Days Inn	803	881-1000	$18	$24
Charleston	Holiday Inn N	803	744-1621	$36	$46
Charleston	Holiday Inn SW	803	556-7100	$33	$47
Charleston	Howard Johnsons	803	524-4148	$31	$36
Charleston	Ramada Inn	803	774-8281	$33	$40
Charleston	Sheraton Inn	803	744-2401	$34	$42
Columbia	Best Western	803	796-9400	$29	$34
Columbia	Carolina Inn	803	799-8200	$42	$48
Columbia	Days Inn	803	736-0000	$23	$27
Columbia	Holiday Inn NW	803	794-9440	$32	$39
Columbia	Howard Johnsons	803	772-7200	$25	$27
Columbia	Quality Inn	803	772-0270	$34	$41
Columbia	Ramada Inn	803	796-2700	$36	$44
Columbia	Vagabond Inn	803	796-6240	$27	$30

Pennsylvania
Bedford Motel/Hotel: Crinaline Courts
(814) 623-9511 S: $18 D: $20
Bedford Motel/Hotel: Holiday Inn
(814) 623-9006 S: $29 D: $36
Bedford Motel/Hotel: Midway
(814) 623-8107 S: $21 D: $26
Bedford Motel/Hotel: Penn Manor
(814) 623-8177 S: $19 D: $25
Bedford Motel/Hotel: Quality Inn
(814) 623-5189 S: $23 D: $28
Bedford Motel/Hotel: Terrace
(814) 623-5111 S: $22 D: $24
Bradley Motel/Hotel: De Soto
(814) 362-3567 S: $20 D: $24
Bradley Motel/Hotel: Holiday House
(814) 362-4511 S: $22 D: $25
Bradley Motel/Hotel: Holiday Inn
(814) 362-4501 S: $32 D: $40
Breezewood Motel/Hotel: Best Western Plaza
(814) 735-4352 S: $20 D: $27
Breezewood Motel/Hotel: Motel 70
(814) 735-4385 S: $16 D: $18

Figure 3.2 Two different ways of structuring the same information at the interface: one makes it much easier to find information than the other. Look at the top screen and: (i) find the price for a double room at the Quality Inn in Columbia; (ii) find the phone number of the Days Inn in Charleston. Then look at the bottom screen and (i) find the price of a double room at the Holiday Inn in Bradley; (ii) find the phone number of the Quality Inn in Bedford. Which took longer to do? In an early study Tullis found that the two screens produced quite different results: it took an average of 3.2 seconds to search the top screen and 5.5 seconds to find the same kind of information in the bottom screen. Why is this so, considering that both displays have the same density of information (31%)? The primary reason is the way the characters are grouped in the display: in the top they are grouped into vertical categories of information (e.g., place, kind of accommodation, phone number, and rates) that have columns of space between them. In the bottom screen the information is bunched up together, making it much harder to search through.

displays. However, it is much harder to find the information in the bottom screen than in the top screen. The reason for this is that the information is very poorly structured in the bottom, making it difficult to find the information. In the top the information has been ordered into meaningful categories with blank spacing between them, making it easier to select the necessary information.

Perception refers to how information is acquired from the environment, via the different sense organs (e.g., eyes, ears, fingers) and transformed into experiences of objects, events, sounds, and tastes (Roth, 1986). It is a complex process, involving other cognitive processes such as memory, attention, and language. Vision is the

DESIGN IMPLICATIONS Attention

- Make information salient when it needs attending to at a given stage of a task.
- Use techniques like animated graphics, color, underlining, ordering of items, sequencing of different information, and spacing of items to achieve this.
- Avoid cluttering the interface with too much information. This especially applies to the use of color, sound and graphics: there is a temptation to use lots of them, resulting in a mishmash of media that is distracting and annoying rather than helping the user attend to relevant information.
- Interfaces that are plain are much easier to use, like the Google search engine (see Figure 3.3). The main reason is that it is much easier for users to find where on the screen to type in their search.

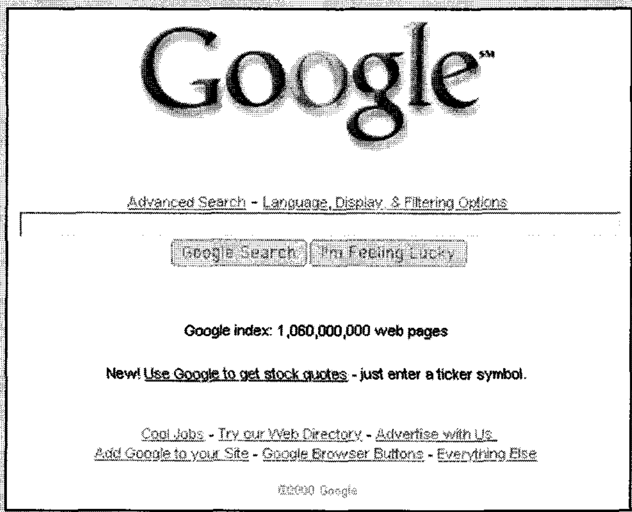

Figure 3.3 Google search engine interface.

most dominant sense for sighted individuals, followed by hearing and touch. With respect to interaction design, it is important to present information in a way that can be readily perceived in the manner intended. For example, there are many ways to design icons. The key is to make them easily distinguishable from one another and to make it simple to recognize what they are intended to represent (not like the ones in Figure 3.4).

Combinations of different media need also to be designed to allow users to recognize the composite information represented in them in the way intended. The use of sound and animation together needs to be coordinated so they happen in a logical sequence. An example of this is the design of lip-synch applications, where the animation of an avatar's or agent's face to make it appear to be talking, must be carefully synchronized with the speech that is emitted. A slight delay between the two can make it difficult and disturbing to perceive what is happening—as sometimes happens when film dubbing gets out of synch. A general design principle is

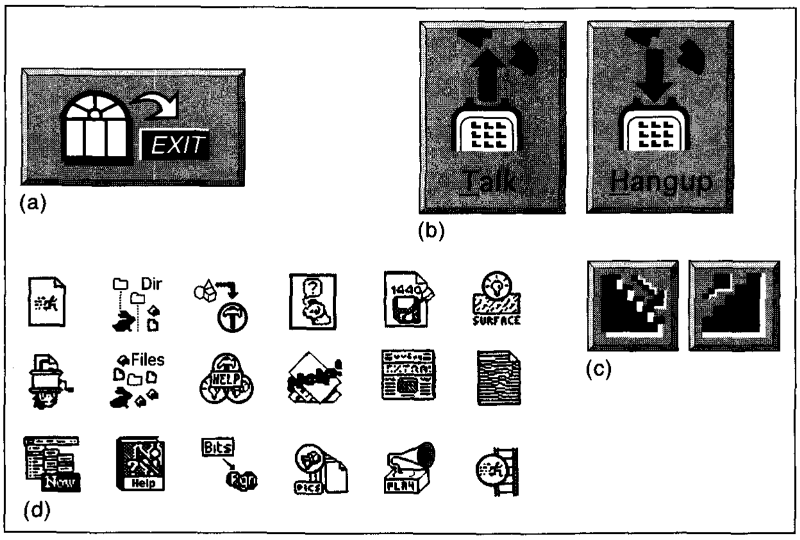

Figure 3.4 Poor icon set. What do you think the icons mean and why are they so bad?

that information needs to be represented in an appropriate form to facilitate the perception and recognition of its underlying meaning.

Memory involves recalling various kinds of knowledge that allow us to act appropriately. It is very versatile, enabling us to do many things. For example, it allows us to recognize someone's face, remember someone's name, recall when we last met them and know what we said to them last. Simply, without memory we would not be able to function.

It is not possible for us to remember everything that we see, hear, taste, smell, or touch, nor would we want to, as our brains would get completely overloaded. A filtering process is used to decide what information gets further processed and memorized. This filtering process, however, is not without its problems. Often we

DESIGN IMPLICATIONS Perception

Representations of information need to be designed to be perceptible and recognizable across different media:

- Icons and other graphical representations should enable users to readily distinguish their meaning.
- Sounds should be audible and distinguishable so users understand what they represent.
- Speech output should enable users to distinguish between the set of spoken words and also be able to understand their meaning.
- Text should be legible and distinguishable from the background (e.g., it is OK to use yellow text on a black or blue background but not on a white or green background).
- Tactile feedback used in virtual environments should allow users to recognize the meaning of the various touch sensations being emulated. The feedback should be distinguishable so that, for example, the sensation of squeezing is represented in a tactile form that is different from the sensation of pushing.

forget things we would dearly love to remember and conversely remember things we would love to forget. For example, we may find it difficult to remember everyday things like people's names and phone numbers or academic knowledge like mathematical formulae. On the other hand, we may effortlessly remember trivia or tunes that cycle endlessly through our heads.

How does this filtering process work? Initially, encoding takes place, determining which information is attended to in the environment and how it is interpreted. The extent to which it takes place affects our ability to recall that information later. The more attention that is paid to something and the more it is processed in terms of thinking about it and comparing it with other knowledge, the more likely it is to be remembered. For example, when learning about a topic it is much better to reflect upon it, carry out exercises, have discussions with others about it, and write notes than just passively read a book or watch a video about it. Thus, how information is interpreted when it is encountered greatly affects how it is represented in memory and how it is used later.

Another factor that affects the extent to which information can be subsequently retrieved is the context in which it is encoded. One outcome is that sometimes it can be difficult for people to recall information that was encoded in a different context from the one they currently are in. Consider the following scenario:

> *You are on a train and someone comes up to you and says hello. You don't recognize him for a few moments but then realize it is one of your neighbors. You are only used to seeing your neighbor in the hallway of your apartment block and seeing him out of context makes him difficult to recognize initially.*

Another well-known memory phenomenon is that people are much better at recognizing things than recalling things. Furthermore, certain kinds of information are easier to recognize than others. In particular, people are very good at recognizing thousands of pictures, even if they have only seen them briefly before.

ACTIVITY 3.1 Try to remember the dates of all the members of your family's and your closest friends' birthdays. How many can you remember? Then try to describe what is on the cover of the last DVD/CD or record you bought. Which is easiest and why?

Comment It is likely that you remembered much better what was on the CD/DVD/record cover (the image, the colors, the title) than the birthdays of your family and friends. People are very good at remembering visual cues about things, for example the color of items, the location of objects (a book being on the top shelf), and marks on an object (e.g., a scratch on a watch, a chip on a cup). In contrast, people find other kinds of information persistently difficult to learn and remember, especially arbitrary material like birthdays and phone numbers.

Instead of requiring users to recall from memory a command name from a possible set of hundreds or even thousands, GUIs provide visually based options that users can browse through until they recognize the operation they want to perform (see Figure 3.5(a) and (b)). Likewise, web browsers provide a facility of bookmarking or saving favorite URLs that have been visited, providing a visual list. This means that users need only recognize a name of a site when scanning through the saved list of URLs.

```
Microsoft Windows 2000 [Version 5.00.2195]
<c> Copyright 1985-1999 Microsoft Corp.

C:\>dir /w
 Volume in drive C has no label
 Volume Serial Number is 07D1-0109

 Directory of C:\

[BACKUP]                  [DELL]                   [DISCOVER]
[I386]                    [WINNT]                  [DRIVERS]
[Documents and settings] [Program Files]           [temp]
[DellUtil]                [DMI]                    [My Music]
[Downloads]               [Palm]                   [Inetpub]
TxE8 - Backup
               1 File(s)          1.367 bytes
              15 Dir(s)  30.522.605.568 bytes free

C:\>cd Documents and settings

C:\Documents and settings>>dir
 Volume in drive C has no label.
 Volume Serial Number is 07D1-0109

 Directory of C:\Documents and settings

09/01/2001  11:49       <DIR>          .
09/01/2001  11:49       <DIR>          ..
09/01/2001  11:49       <DIR>          All Users
09/01/2001  12:04       <DIR>          Administrator
               0 File<s>             0 bytes
               4Dir<s>   30,522,605,568 bytes free

C:\Documents and settings>cd Administrator

C:\Documents and Settings\Administrator>dir
 Volume in drive C has no label..
 Volume Serial Number is 07D1-0109

 Directory of C:\Documents and settings\Administrator

09/01/2001  12:04       <DIR>          .
09/01/2001  12:04       <DIR>          ..
09/01/2001  11:49       <DIR>          Start Menu
09/01/2001  11:49       <DIR>          My Documents
09/01/2001  11:49       <DIR>          Favorites
09/01/2001  11:49       <DIR>          Desktop
24/01/2001  17:16       <DIR>          Abisuite
               0 File<s>             0 bytes
               7Dir<s>   30,522,605,568 bytes free

C:\Documents and settings\Administrator>cd My Documents

C:\Documents and settings\Administrator\My Documents>
```

Figure 3.5(a) A DOS-based interface, requiring the user to type in commands.

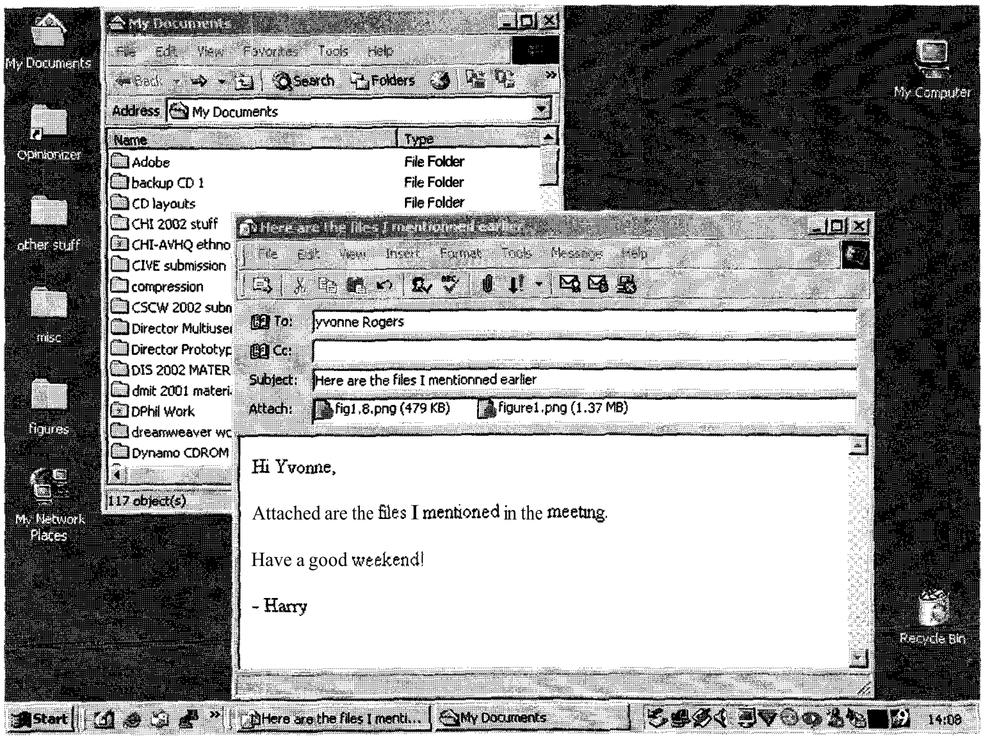

Figure 3.5(b) A Windows-based interface, with menus, icons, and buttons.

ACTIVITY 3.2 What strategies do you use to help you remember things?

Comment People often write down what they need to remember on a piece of paper. They also ask others to remind them. Another approach is to use various mental strategies, like mnemonics. A mnemonic involves taking the first letters of a set of words in a phrase or set of concepts and using them to make a more memorable phrase, often using bizarre and idiosyncratic connections. For example, some people have problems working out where east is in relation to west and vice versa (i.e., is it to the left or right). A mnemonic to help figure this out is to take the first letters of the four main points of the compass and then use them in the phrase "Never Eat Shredded Wheat" mentally recited in a clockwise sequence.

A growing problem for computer users is file management. The number of documents created, images and videoclips downloaded, emails and attachments saved, URLs bookmarked, and so on increases every day. A major problem is finding them again. Naming is the most common means of encoding them, but trying to remember a name of a file you created some time back can be very difficult, especially if there are tens of thousands of named files. How might such a process be facilitated, bearing in mind people's memory abilities? Mark Lansdale, a British psychologist, has been researching this problem of information retrieval for many

BOX 3.1 The Problem with the Magical Number 7 Plus or Minus 2

Perhaps the best known finding in psychology (certainly the one that nearly all students remember many years after they have finished their studies) is George Miller's (1956) theory that 7±2 chunks of information can be held in short-term memory at any one time. By short-term memory he meant a memory store in which information was assumed to be processed when first perceived. By chunks he meant a range of items like numbers, letters, or words. According to Miller's theory, therefore, people's immediate memory capacity is very limited. They are able to remember only a few words or numbers that they have heard or seen. If you are not familiar with this phenomenon, try out the following exercise: read the first list below (or get someone to read it to you), cover it up, and then try to recall as many of the items as possible. Repeat this for the other lists.

- 3, 12, 6, 20, 9, 4, 0, 1, 19, 8, 97, 13, 84
- cat, house, paper, laugh, people, red, yes, number, shadow, broom, rain, plant, lamp, chocolate, radio, one, coin, jet
- t, k, s, y, r, q, x, p, a, z, l, b, m, e

How many did you correctly remember for each list? Between 5 and 9, as suggested by Miller's theory?

Chunks can also be combined items that are meaningful. For example, it is possible to remember the same number of two-word phrases like hot chocolate, banana split, cream cracker, rock music, cheddar cheese, leather belt, laser printer, tree fern, fluffy duckling, cold rain. When these are all muddled up (e.g., split belt, fern crackers, banana laser, printer cream, cheddar tree, rain duckling, hot rock), however, *it is much harder to remember* as many chunks. This is mainly because the first set contains all meaningful two-word phrases that have been heard before and require less time to be processed in short-term memory, whereas the second set are completely novel phrases that don't exist in the real world. You need to spend time linking the two parts of the phrase together while trying to memorize them. This takes more time and effort to achieve. Of course it is possible to do if you have time to spend rehearsing them, but if you are asked to do it having heard them only once in quick succession, it is most likely you will remember only a few.

You may be thinking by now, "OK, this is interesting, but what has it got to do with interaction design?" Well not only does this 50-year-old theory have a special place in psychology, it has also made a big impression in HCI. Unfortunately, however, for the wrong reasons. Many designers have heard or read about this phenomenon and think, ah, here is a bit of psychology I can usefully apply to interface design. Would you agree with them? If so, how might people's ability to only remember 7±2 chunks that they have just read or heard be usefully applied to interaction design?

According to a survey by Bob Bailey (2000), several designers have been led to believe the following guidelines and have even created interfaces based on them:

- Have only seven options on a menu.
- Display only seven icons on a menu bar.
- Never have more than seven bullets in a list.
- Place only seven tabs at the top of a website page.
- Place only seven items on a pull-down menu.

All of these are wrong. Why? The simple reason is that these are all items that can be scanned and rescanned visually and hence do *not* have to be recalled from short-term memory. They don't just flash up on the screen and disappear, requiring the user to remember them before deciding which one to select. If you were asked to find an item of food most people crave in the set of single words listed above, would you have any problem? No, you would just scan the list until you recognized the one (chocolate) that matched the task and then select it—just as people do when interacting with menus, lists, and tabs—regardless of whether they comprise three or 30 items. What the users are required to do here is not remember as many items as possible, having only heard or seen them once in a sequence, but instead *scan* through a set of items until they *recognize* the one they want. Quite a different task. Furthermore, there is much more useful psychological research that can be profitably applied to interaction design.

years. He suggests that it is profitable to view this process as involving two memory processes: recall-directed, followed by recognition-based scanning. The first refers to using memorized information about the required file to get as close to it as possible. The more exact this is, the more success the user will have in tracking down the desired file. The second happens when recall has failed to produce what a user wants and so requires reading through directories of files.

To illustrate the difference between these two processes, consider the following scenario: a user is trying to access a couple of websites visited the day before that compared the selling price of cars offered by different dealers. The user is able to recall the name of one website: "alwaysthecheapest.com". She types this in and the website appears. This is an example of successful recall-directed memory. However, the user is unable to remember the name of the second one. She vaguely remembers it was something like 'autobargains.com'; but typing this in proves unsuccessful. Instead, she switches to scanning her bookmarks/favorites, going to the list of most recent ones saved. She notices two or three URLs that could be the one desired, and on the second attempt she finds the website she is looking for. In this situation, the user initially tries recall-directed memory and when this fails, adopts the second strategy of recognition-based scanning—which takes longer but eventually results in success.

Lansdale proposes that file management systems should be designed to optimize both kinds of memory processes. In particular, systems should be developed that let users use whatever memory they have to limit the area being searched and then represent the information in this area of the interface so as to maximally assist them in finding what they need. Based on this theory, he has developed a prototype system called MEMOIRS that aims at improving users' recall of information they had encoded so as to make it easier to recall later (Lansdale and Edmunds, 1992). The system was designed to be flexible, providing the user with a range of ways of encoding documents mnemonically, including time stamping (see Figure 3.6), flagging, and attribution (e.g., color, text, icon, sound or image).

More flexible ways of helping users track down the files they want are now beginning to be introduced as part of commercial applications. For example, various search and find tools, like Apple's Sherlock, have been designed to enable the user to type a full or partial name or phrase that the system then tries to match by listing all the files it identifies containing the requested name/phrase. This method, however, is still quite limited, in that it allows users to encode and retrieve files using only alphanumericals.

DESIGN IMPLICATIONS Memory

- Do not overload users' memories with complicated procedures for carrying out tasks.
- Design interfaces that promote *recognition* rather than *recall* by using menus, icons, and consistently placed objects.
- Provide users with a variety of ways of encoding electronic information (e.g., files, emails, images) to help them remember where they have stored them, through the use of color, flagging, time stamping, icons, etc.

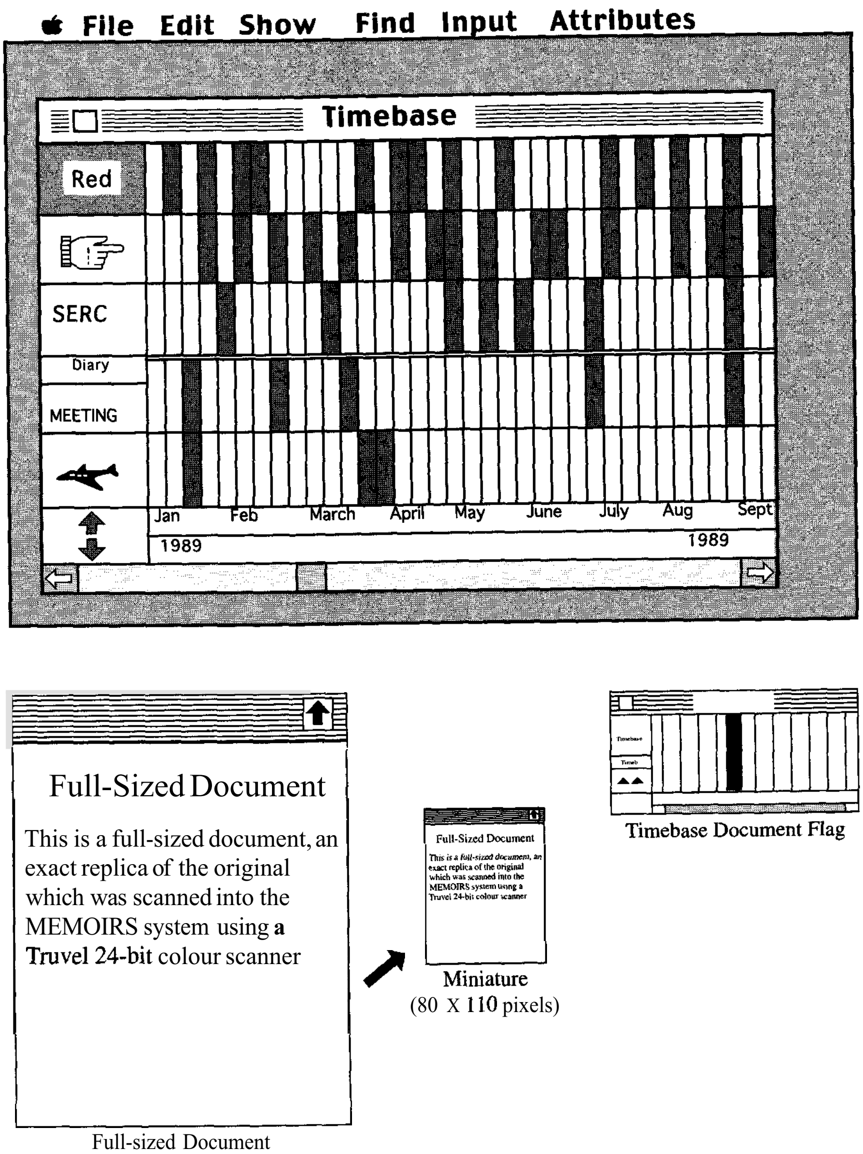

Figure 3.6 Memoirs tool.

BOX 3.2 A Case of Too Much Memory Load?

Phone banking has become increasingly popular in the last few years. It allows customers to carry out financial transactions, such as paying bills and checking the balance of their accounts, at their convenience. One of the problems confronting banks that provide this facility, however, is how to manage security concerns. Anyone can phone up a bank and pretend to be someone else. How do the banks prevent fraudulent transactions?

One solution has been to develop rigorous security measures whereby customers must provide various pieces of information before gaining access to their accounts. Typically, these include providing the answers to a combination of the following:

- their zip code or post code
- their mother's maiden name
- their birthplace
- the last school they attended
- the first school they attended
- a password of between 5 and 10 letters
- a *memorable* address (not their home)
- a *memorable* date (not their birthday)

Many of these are relatively easy to remember and recall as they are very familiar. But consider the last two. How easy is it for someone to come up with such *memorable* information and then be able to recall it readily? Perhaps the customer can give the address and birthday of another member of their family as a memorable address and date. But what about the request for a password? Suppose a customer selects the word "interaction" as a password—fairly easy to remember. The problem is that the bank operators do not ask for the full password, because of the danger that someone in the vicinity might overhear and write it down. Instead they are instructed to ask the customer to provide specific letters from it, like the 7th followed by the 5th. However, such information does not spring readily to mind. Instead, it requires mentally counting each letter of the password until the desired one is reached. How long does it take you to determine the 7th letter of the password "interaction"? How did you do it?

To make things harder, banks also randomize the questions they ask. Again, this is to prevent someone who might be overhearing from memorizing the sequence of information. However, it also means that the customers themselves cannot learn the sequence of information required, meaning they have to generate different information every time they call up the bank.

This requirement to remember and recall such information puts a big memory load on customers. Some people find such a procedure quite nerve-wracking and are prone to forget certain pieces of information. As a coping strategy they write down their details on a sheet of paper. Having such an external representation at hand makes it much easier for them to read off the necessary information rather than having to recall it from memory. However, it also makes them vulnerable to the very fraud the banks were trying to prevent, should anyone else get hold of that piece of paper!

ACTIVITY 3.3 How else might banks solve the problem of providing a secure system while making the memory load relatively easy for people wanting to use phone banking? How does phone banking compare with online banking?

Comment An alternative approach is to provide the customers with a PIN number (it could be the same as that of their ATM card) and ask them to key this in on their phone keypad, followed by asking one or two questions like their zip or post code, as a backup. Online banking has similar security risks to phone banking and hence this requires a number of security measures to be enforced. These include that the user sets up a nickname and a password. For example, some banks require typing in three randomly selected letters from a password each time the user logs on. This is harder to do online than when asked over the phone, mainly because it interferes with the normally highly automated process of typing in a password. You really have to think about what letters and numbers are in your password; for example, has it got two letter f's after the number 6, or just one?

Learning can be considered in terms of (i) how to use a computer-based application or (ii) using a computer-based application to understand a given topic. Jack Carroll (1990) and his colleagues have written extensively about how to design interfaces to help learners develop computer-based skills. A main observation is that people find it very hard to learn by following sets of instructions in a manual. Instead, they much prefer to "learn through doing." GUIs and direct manipulation interfaces are good environments for supporting this kind of learning by supporting exploratory interaction and importantly allowing users to "undo" their actions, i.e., return to a previous state if they make a mistake by clicking on the wrong option. Carroll has also suggested that another way of helping learners is by using a "training-wheels" approach. This involves restricting the possible functions that can be carried out by a novice to the basics and then extending these as the novice becomes more experienced. The underlying rationale is to make initial learning more tractable, helping the learner focus on simple operations before moving on to more complex ones.

There have also been numerous attempts to harness the capabilities of different technologies to help learners understand topics. One of the main benefits of interactive technologies, such as web-based, multimedia, and virtual reality, is that they provide alternative ways of representing and interacting with information that are not possible with traditional technologies (e.g., books, video). In so doing, they have the potential of offering learners the ability to explore ideas and concepts in different ways.

ACTIVITY 3.4 Ask a grandparent, child, or other person who has not used a cell phone before to make and answer a call using it. What is striking about their behavior?

Comment First-time users often try to apply their understanding of a land-line phone to operating a cell phone. However, there are marked differences in the way the two phones operate, even for the simplest of tasks, like making a call. First, the power has to be switched on when using a cell phone, by pressing a button (but not so with land-line phones), then the number has to be keyed in, including at all times the area code (in the UK), even if the callee is in the same area (but not so with land-lines), and finally the "make a call" button must be pressed (but not so with land-line phones). First-time users may intuitively know how to switch the phone on but not know which key to hit, or that it has to be held down for a couple of seconds. They may also forget to key in the area code if they are in the same area as the person they are calling, and to press the "make a call" key. They may also forget to press the "end a call" button (this is achieved through putting the receiver down with a land-line phone). Likewise, when answering a call, the first-time user may forget to press the "accept a call" button or not know which one to press. These additional actions are quick to learn, once the user understands the need to explicitly instruct the cell phone when they want to make, accept, or end a call.

Reading, speaking and listening: these three forms of language processing have both similar and different properties. One similarity is that the meaning of

BOX 3.3 Learning the "Difficult Stuff" through Interactive Multimedia: the Role of Dynalinking

Children (and adults) often have problems learning the difficult stuff—by this we mean mathematical formulae, notations, laws of physics, and other abstract concepts. One of the main reasons is that they find it difficult to relate their concrete experiences of the physical world with these higher-level abstractions. Research has shown, however, that it is possible to facilitate this kind of learning through the use of interactive multimedia. In particular, different representations of the same process (e.g., a graph, a formula, a sound, a simulation) can be displayed and interacted with in ways that make their relationship with each other more explicit to the learner. This process of linking and manipulating multimedia representations at the interface is called dynalinking (Rogers and Scaife, 1998).

An example where we have found dynalinking beneficial is in helping children and students learn ecological concepts (e.g., food webs, carbon cycles, and energy). In one of our projects, we built a simple ecosystem of a pond using multimedia. The concrete simulation showed various organisms swimming and moving around and occasionally an event where one would eat another (e.g., a snail eating the weed). This was annotated and accompanied by various eating sounds (e.g., chomping) to attract the children's attention. The children could also interact with the simulation. When an organism was clicked on, it would say what it was and what it ate (e.g., "I'm a weed. I make my own food").

The simulation was dynalinked with other abstract representations of the pond ecosystem. One of these was a food web diagram (See Figure 3.7 in Color Plate 4). The children were encouraged to interact with the interlinked diagrams in various ways and to observe what happened in the concrete simulation when something was changed in the diagram and vice versa. Our study showed that children enjoyed interacting with the simulation and diagrams and, importantly, that they understood much better the purpose of the abstract diagrams and how to use them to reason about the ecosystem.

Dynalinking is a powerful form of interaction and can be used in a range of domains to explicitly show relationships among multiple dimensions, especially when the information to be understood or learned is complex. For example, it can be useful for domains like economic forecasting, molecular modeling, and statistical analyses.

sentences or phrases is the same regardless of the mode in which it is conveyed. For example, the sentence "Computers are a wonderful invention" essentially has the same meaning whether one reads it, speaks it, or hears it. However, the ease with which people can read, listen, or speak differs depending on the person, task, and context. For example, many people find listening much easier than reading. Specific differences between the three modes include:

- Written language is permanent while listening is transient. It is possible to reread information if not understood the first time round. This is not possible with spoken information that is being broadcast.

DESIGN IMPLICATIONS Learning

- Design interfaces that encourage exploration.
- Design interfaces that constrain and guide users to select appropriate actions.
- Dynamically link representations and abstractions that need to be learned.

- Reading can be quicker than speaking or listening, as written text can be rapidly scanned in ways not possible when listening to serially presented spoken words.
- Listening requires less cognitive effort than reading or speaking. Children, especially, often prefer to listen to narratives provided in multimedia or web-based learning material than to read the equivalent text online.
- Written language tends to be grammatical while spoken language is often ungrammatical. For example, people often start a sentence and stop in mid-sentence, letting someone else start speaking.
- There are marked differences between people in their ability to use language. Some people prefer reading to listening, while others prefer listening. Likewise, some people prefer speaking to writing and vice versa.
- Dyslexics have difficulties understanding and recognizing written words, making it hard for them to write grammatical sentences and spell correctly.
- People who are hard of hearing or hard of seeing are also restricted in the way they can process language.

Many applications have been developed either to capitalize on people's reading, writing and listening skills, or to support or replace them where they lack or have difficulty with them. These include:

- interactive books and web-based material that help people to read or learn foreign languages
- speech-recognition systems that allow users to provide instructions via spoken commands (e.g., word-processing dictation, home control devices that respond to vocalized requests)
- speech-output systems that use artificially generated speech (e.g., written-text-to-speech systems for the blind)
- natural-language systems that enable users to type in questions and give text-based responses (e.g., Ask Jeeves search engine)
- cognitive aids that help people who find it difficult to read, write, and speak. A number of special interfaces have been developed for people who have problems with reading, writing, and speaking (e.g., see Edwards, 1992).
- various input and output devices that allow people with various disabilities to have access to the web and use word processors and other software packages

Helen Petrie and her team at the Sensory Disabilities Research Lab in the UK have been developing various interaction techniques to allow blind people to access the web and other graphical representations, through the use of auditory navigation and tactile diagrams.

Problem-solving, planning, reasoning and decision-making are all cognitive processes involving reflective cognition. They include thinking about what to do, what the options are, and what the consequences might be of carrying out a given action. They often involve conscious processes (being aware of what one is thinking

DESIGN IMPLICATIONS Reading, Speaking and Listening

- Keep the length of speech-based menus and instructions to a minimum. Research has shown that people find it hard to follow spoken menus with more than three or four options. Likewise, they are bad at remembering sets of instructions and directions that have more than a few parts.
- Accentuate the intonation of artificially generated speech voices, as they are harder to understand than human voices.
- Provide opportunities for making text large on a screen, without affecting the formatting, for people who find it hard to read small text.

about), discussion with others (or oneself), and the use of various kinds of artifacts, (e.g., maps, books, and pen and paper). For example, when planning the best route to get somewhere, say a foreign city, we may ask others, use a map, get instructions from the web, or a combination of these. Reasoning also involves working through different scenarios and deciding which is the best option or solution to a given problem. In the route-planning activity we may be aware of alternative routes and reason through the advantages and disadvantages of each route before deciding on the best one. Many a family argument has come about because one member thinks he or she knows the best route while another thinks otherwise.

Comparing different sources of information is also common practice when seeking information on the web. For example, just as people will phone around for a range of quotes, so too, will they use different search engines to find sites that give the best deal or best information. If people have knowledge of the pros and cons of different search engines, they may also select different ones for different kinds of queries. For example, a student may use a more academically oriented one when looking for information for writing an essay, and a more commercially based one when trying to find out what's happening in town.

The extent to which people engage in the various forms of reflective cognition depends on their level of experience with a domain, application, or skill. Novices tend to have limited knowledge and will often make assumptions about what to do using other knowledge about similar situations. They tend to act by trial and error, exploring and experimenting with ways of doing things. As a result they may start off being slow, making errors and generally being inefficient. They may also act irrationally, following their superstitions and not thinking ahead to the consequences of their actions. In contrast, experts have much more knowledge and experience and are able to select optimal strategies for carrying out their tasks. They are likely to be able to think ahead more, considering what the consequences might be of opting for a particular move or solution (as do expert chess players).

DESIGN IMPLICATION Problem-Solving, Planning, Reasoning and Decision-Making

- Provide additional hidden information that is easy to access for users who wish to understand more about how to carry out an activity more effectively (e.g., web searching).

3.3 Applying knowledge from the physical world to the digital world

As well as understanding the various cognitive processes that users engage in when interacting with systems, it is also useful to understand the way people cope with the demands of everyday life. A well known approach to applying knowledge about everyday psychology to interaction design is to emulate, in the digital world, the strategies and methods people commonly use in the physical world. An assumption is that if these work well in the physical world, why shouldn't they also work well in the digital world? In certain situations, this approach seems like a good idea. Examples of applications that have been built following this approach include electronic post-it notes in the form of "stickies," electronic "to-do" lists, and email reminders of meetings and other events about to take place. The stickies application displays different colored notes on the desktop in which text can be inserted, deleted, annotated, and shuffled around, enabling people to use them to remind themselves of what they need to do—analogous to the kinds of externalizing they do when using paper stickies. Moreover, a benefit is that electronic stickies are more durable than paper ones—they don't get lost or fall off the objects they are stuck to, but stay on the desktop until explicitly deleted.

In other situations, however, the simple emulation approach can turn out to be counter-productive, forcing users to do things in bizarre, inefficient, or inappropriate ways. This can happen when the activity being emulated is more complex than is assumed, resulting in much of it being oversimplified and not supported effectively. Designers may notice something salient that people do in the physical world and then fall into the trap of trying to copy it in the electronic world without thinking through how and whether it will work in the new context (remember the poor design of the virtual calculator based on the physical calculator described in the previous chapter).

Consider the following classic study of real-world behavior. Ask yourself, first, whether it is useful to emulate at the interface, and second, how it could be extended as an interactive application.

Tom Malone (1983) carried out a study of the "natural history" of physical offices. He interviewed people and studied their offices, paying particular attention to their filing methods and how they organized their papers. One of his findings was that whether people have messy offices or tidy offices may be more significant than people realize. Messy offices were seen as being chaotic with piles of papers everywhere and little organization. Tidy offices, on the other hand, were seen as being well organized with good use of a filing system. In analyzing these two types of offices, Malone suggested what they reveal in terms of the underlying cognitive behaviors of the occupants. One of his observations was that messy offices may appear chaotic but in reality often reflect a coping strategy by the person: documents are left lying around in obvious places to act as reminders that something has to be done with them. This observation suggests that using piles is a fundamental strategy, regardless of whether you are a chaotic or orderly person.

Such observations about people's coping strategies in the physical world bring to mind an immediate design implication about how to support electronic file management: to capitalize on the "pile" phenomenon by trying to emulate it in the electronic world. Why not let people arrange their electronic files into piles as they do with paper files? The danger of doing this is that it could heavily constrain the way people manage their files, when in fact there may be far more effective and flexible ways of filing in the electronic world. Mark Lansdale (1988) points out how introducing unstructured piles of electronic documents on a desktop would be counterproductive, in the same way as building planes to flap their wings in the way birds do (someone seriously thought of doing this).

But there may be benefits of emulating the pile phenomenon by using it as a kind of interface metaphor that is extended to offer other functionality. How might this be achieved? A group of interface designers at Apple Computer (Mandler et al., 1992) tackled this problem by adopting the philosophy that they were going to build an application that went beyond physical-world capabilities, providing new functionality that only the computer could provide and that enhanced the interface. To begin their design, they carried out a detailed study of office behavior and analyzed the many ways piles are created and used. They also examined how people use the default hierarchical file-management systems that computer operating systems provide. Having a detailed understanding of both enabled them to create a conceptual model for the new functionality—which was to provide various interactive organizational elements based around the notion of using piles. These included providing the user with the means of creating, ordering, and visualizing piles of files. Files could also be encoded using various external cues, including date and color. New functionality that could not be achieved with physical files included the provision of a scripting facility, enabling files in piles to be ordered in relation to these cues (see Figure 3.8).

Emulating real-world activity at the interface can be a powerful design strategy, provided that new functionality is incorporated that extends or supports the users in their tasks in ways not possible in the physical world. The key is really to understand the nature of the problem being addressed in the electronic world in relation to the various coping and externalizing strategies people have developed to deal with the physical world.

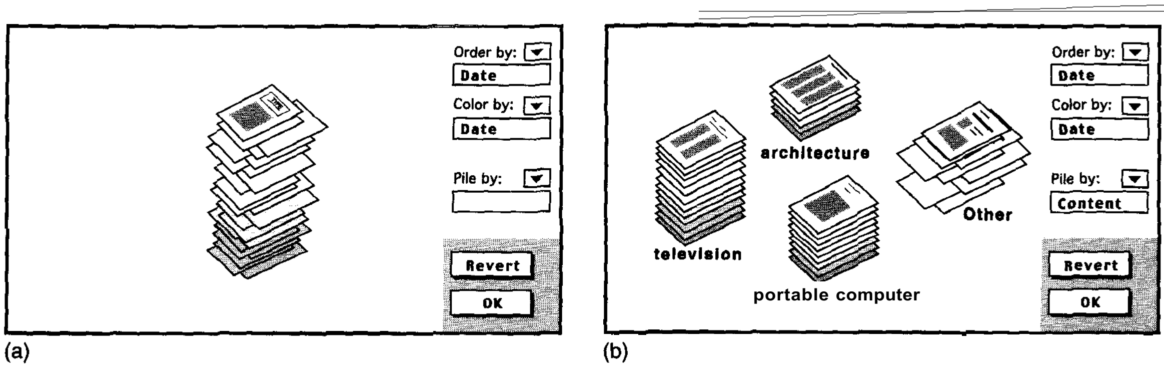

Figure 3.8 The pile metaphor as it appears at the interface.

3.4 Conceptual frameworks for cognition

In the previous section we described the pros and cons of applying knowledge of people's coping strategies in the physical world to the digital world. Another approach is to apply theories and conceptual frameworks to interaction design. In this section we examine three of these approaches, which each have a different perspective on cognition:

- mental models
- information processing
- external cognition

3.4.1 Mental models

In Chapter 2 we pointed out that a successful system is one based on a conceptual model that enables users to readily learn a system and use it effectively. What happens when people are learning and using a system is that they develop knowledge of how to use the system and, to a lesser extent, how the system works. These two kinds of knowledge are often referred to as a user's mental model.

Having developed a mental model of an interactive product, it is assumed that people will use it to make inferences about how to carry out tasks when using the interactive product. Mental models are also used to fathom what to do when something unexpected happens with a system and when encountering unfamiliar systems. The more someone learns about a system and how it functions, the more their mental model develops. For example, TV engineers have a "deep" mental model of how TVs work that allows them to work out how to fix them. In contrast.

an average citizen is likely to have a reasonably good mental model of how to operate a TV but a "shallow" mental model of how it works.

Within cognitive psychology, mental models have been postulated as internal constructions of some aspect of the external world that are manipulated enabling predictions and inferences to be made (Craik, 1943). This process is thought to involve the "fleshing out" and the "running" of a mental model (Johnson-Laird, 1983). This can involve both unconscious and conscious mental processes, where images and analogies are activated.

ACTIVITY 3.5 To illustrate how we use mental models in our everyday reasoning, imagine the following two scenarios:

(a) You arrive home from a holiday on a cold winter's night to a cold house. You have a small baby and you need to get the house warm as quickly as possible. Your house is centrally heated. Do you set the thermostat as high as possible or turn it to the desired temperature (e.g. 70°F)?

(b) You arrive home from being out all night, starving hungry. You look in the fridge and find all that is left is an uncooked pizza. The instructions on the packet say heat the oven to 375°F and then place the pizza in the oven for 20 minutes. Your oven is electric. How do you heat it up? Do you turn it to the specified **temperature** or **higher?**

Comment Most people when asked the first question imagine the scenario in terms of what they would do in their own house and choose the first option. When asked why, a typical explanation that is given is that setting the temperature to be as high as possible increases the rate at which the room warms up. While many people may believe this, it is incorrect. Thermostats work by switching on the-heat and keeping it going at a constant speed until the desired temperature set is reached, at which point they cut out. They cannot control the rate at which heat is given out from a heating system. Left at a given setting, thermostats will turn the heat on and off as necessary to maintain the desired temperature.

When asked the second question, most people say they would turn the oven to the specified temperature and put the pizza in when they think it is at the desired temperature. Some people answer that they would turn the oven to a higher temperature in order to warm it up more quickly. Electric ovens work on the same principle as central heating and so turning the heat up higher will not warm it up any quicker. There is also the problem of the pizza burning if the oven is too hot!

Why do people use erroneous mental models? It seems that in the above scenarios, they are running a mental model based on a general valve theory of the way something works (Kempton, 1986). This assumes the underlying principle of "more is more": the more you turn or push something, the more it causes the desired effect. This principle holds for a range of physical devices, such as taps and radio controls, where the more you turn them, the more water or volume is given. However, it does not hold for thermostats, which instead function based on the principle of an on-off switch. What seems to happen is that in everyday life people develop a core set of abstractions about how things work, and apply these to a range of devices, irrespective of whether they are appropriate.

Using incorrect mental models to guide behavior is surprisingly common. Just watch people at a pedestrian crossing or waiting for an elevator (lift). How many times do they press the button? A lot of people will press it at least twice. When asked why, a common reason given is that they think it will make the lights change faster or ensure the elevator arrives. This seems to be another example of following the "more is more" philosophy: it is believed that the more times you press the button, the more likely it is to result in the desired effect.

Another common example of an erroneous mental model is what people do when the cursor freezes on their computer screen. Most people will bash away at all manner of keys in the vain hope that this will make it work again. However, ask them how this will help and their explanations are rather vague. The same is true when the TV starts acting up: a typical response is to hit the top of the box repeatedly with a bare hand or a rolled-up newspaper. Again, ask people why and their reasoning about how this behavior will help solve the problem is rather lacking.

The more one observes the way people interact with and behave towards interactive devices, the more one realizes just how strange their behavior can get—especially when the device doesn't work properly and they don't know what to do. Indeed, research has shown that people's mental models of the way interactive devices work is poor, often being incomplete, easily confusable, based on inappropriate analogies, and superstition (Norman, 1983). Not having appropriate mental models available to guide their behavior is what causes people to become very frustrated—often resulting in stereotypical "venting" behavior like those described above.

On the other hand, if people could develop better mental models of interactive systems, they would be in a better position to know how to carry out their tasks efficiently and what to do if the system started acting up. Ideally, they should be able to develop a mental model that matches the conceptual model developed by the designer. But how can you help users to accomplish this? One suggestion is to educate them better. However, many people are resistant to spending much time learning about how things work, especially if it involves reading manuals and other documentation. An alternative proposal is to design systems to be more transparent, so that they are easier to understand. This doesn't mean literally revealing the guts of the system (cf. the way some phone handsets—see Figure 3.9 on Color Plate 4—and iMacs are made of transparent plastic to reveal the colorful electronic circuitry inside), but requires developing an easy-to-understand system image (see Chapter 2 for explanation of this term in relation to conceptual models). Specifically, this involves providing:

- useful feedback in response to user input
- easy-to-understand and intuitive ways of interacting with the system

In addition, it requires providing the right kind and level of information, in the form of:

- clear and easy-to-follow instructions
- appropriate online help and tutorials
- context-sensitive guidance for users, set at their level of experience, explaining how to proceed when they are not sure what to do at a given stage of a task.

DILEMMA How Much Transparency?

How much and what kind of transparency do you think a designer should provide in an interactive product? This is not a straightforward question to answer and depends a lot on the requirements of the targeted user groups. Some users simply want to get on with their tasks and don't want to have to learn about how the thing they are using works. In this situation, the system should be designed to make it obvious what to do and how to use it. For example, most cell-phone users want a simple "plug-and-play" type interface, where it is straightforward how to carry out functions like saving an address, text messaging, and making a call. Functions that require too much learning can be off-putting. Users simply won't bother to make the extra effort, meaning that many of the functions provided are never used. Other users like to understand how the device they are using works, in order to make informed decisions about how to carry out their tasks, especially if there are numerous ways of doing something. Some search engines have been designed with this in mind: they provide background information on how they work and how to improve one's searching techniques (see Figure 3.10).

Thus, the extent to which designers should provide extensive information about how to use a system and how it works, as part of the system image, needs to be appraised in terms of what different people want to know and how much they are prepared to learn.

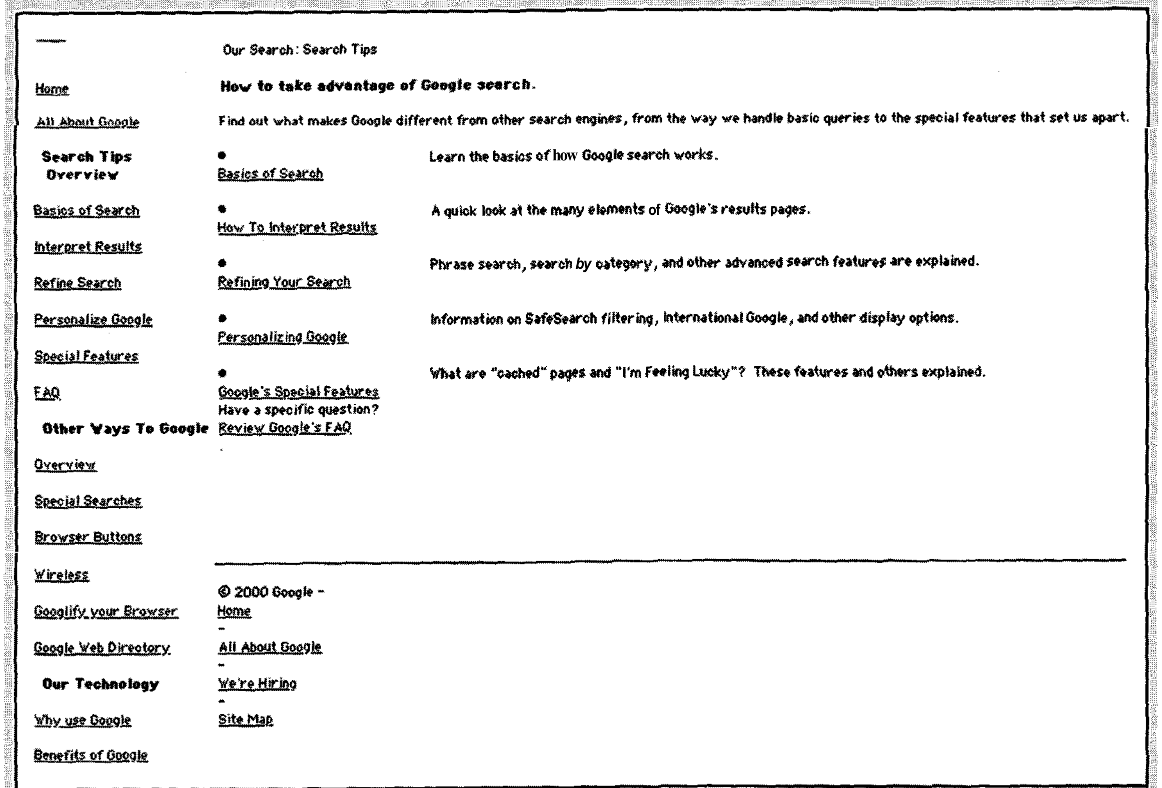

Figure 3.10 The Google search engine, which provides extensive information about how to make your searching strategy more effective.

3.4.2 Information processing

Another approach to conceptualizing how the mind works has been to use metaphors and analogies (see also Chapter 2). A number of comparisons have been made, including conceptualizing the mind as a reservoir, a telephone network, and a digital computer. One prevalent metaphor from cognitive psychology is the idea that the mind is an information processor. Information is thought to enter and exit the mind through a series of ordered processing stages (see Figure 3.11). Within these stages, various processes are assumed to act upon mental representations. Processes include comparing and matching. Mental representations are assumed to comprise images, mental models, rules, and other forms of knowledge.

The information processing model provides a basis from which to make predictions about human performance. Hypotheses can be made about how long someone will take to perceive and respond to a stimulus (also known as reaction time) and what bottlenecks occur if a person is overloaded with too much information. The best known approach is the human processor model, which models the cognitive processes of a user interacting with a computer (Card et al., 1983). Based on the information processing model, cognition is conceptualized as a series of processing stages, where perceptual, cognitive, and motor processors are organized in relation to one another (see Figure 3.12). The model predicts which cognitive processes are involved when a user interacts with a computer, enabling calculations to be made of how long a user will take to carry out various tasks. This can be very useful when comparing different interfaces. For example, it has been used to compare how well different word processors support a range of editing tasks.

The information processing approach is based on modeling mental activities that happen *exclusively* inside the head. However, most cognitive activities involve people interacting with external kinds of representations, like books, documents, and computers—not to mention one another. For example, when we go home from wherever we have been we do not need to remember the details of the route because we rely on cues in the environment (e.g., we know to turn left at the red house, right when the road comes to a T-junction, and so on). Similarly, when we are at home we do not have to remember where everything is because information is "out there." We decide what to eat and drink by scanning the items in the fridge, find out whether any messages have been left by glancing at the answering machine to see if there is a flashing light, and so on. To what extent, therefore, can we say that information processing models are truly representative of everyday cognitive activities? Do they adequately account for cognition as it happens in the real world and, specifically, how people interact with computers and other interactive devices?

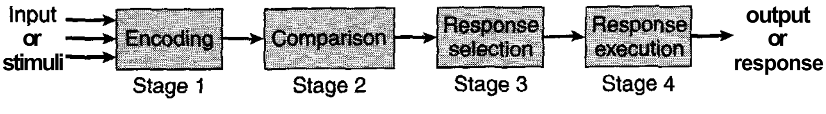

Figure 3.11 Human information processing model.

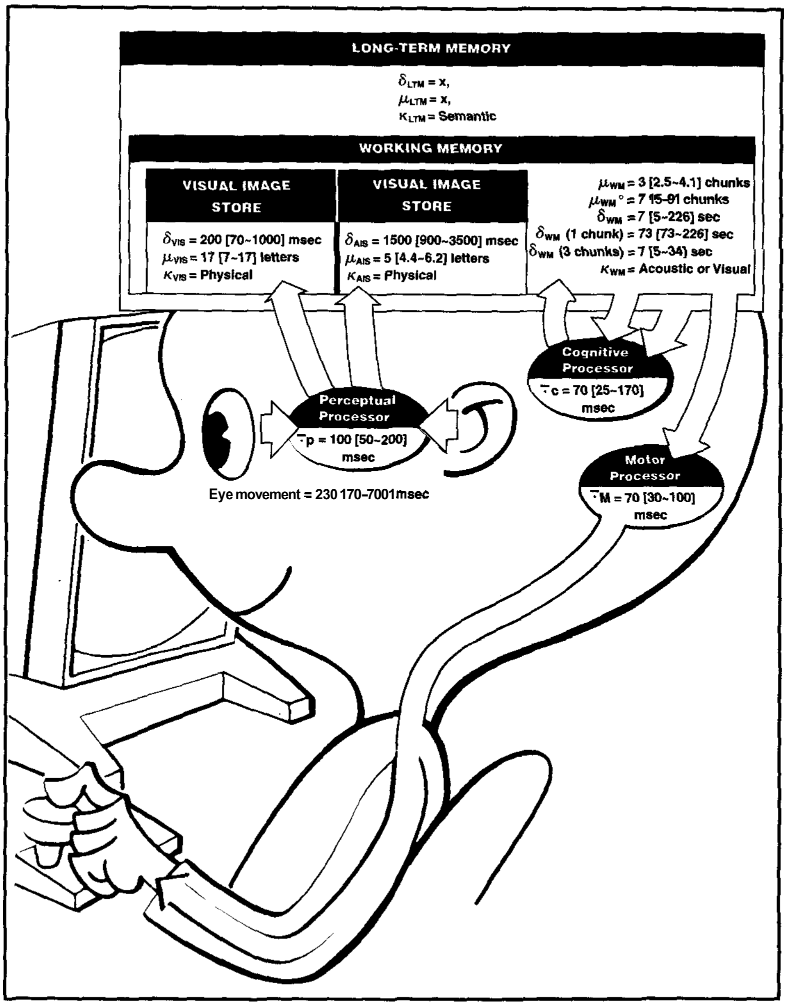

Figure 3.12 The *human processor model.*

Several researchers have argued that existing information processing approaches are too impoverished:

> *The traditional approach to the study of cognition is to look at the pure intellect, isolated from distractions and from artificial aids. Experiments are performed in closed, isolated rooms, with a minimum of distracting lights or sounds, no other people to assist with the task, and no aids to memory or thought. The tasks are arbitrary ones, invented by the researcher. Model builders build simulations and descriptions of these isolated situations. The theoretical analyses are self-contained little structures, isolated from the world, isolated from any other knowledge or abilities* of the *person. (Norman, 1990, p. 5)*

Instead, there has been an increasing trend to study cognitive activities in the context in which they occur, analyzing cognition as it happens "in the wild"

(Hutchins, 1995). A central goal has been to look at how structures in the environment can both aid human cognition and reduce cognitive load. A number of alternative frameworks have been proposed, including external cognition and distributed cognition. In this chapter, we look at the ideas behind external cognition—which has focused most on how to inform interaction design (distributed cognition is described in the next chapter).

3.4.3 External cognition

People interact with or create information through using a variety of external representations, e.g., books, multimedia, newspapers, web pages, maps, diagrams, notes, drawings, and so on. Furthermore, an impressive range of tools has been developed throughout history to aid cognition, including pens, calculators, and computer-based technologies. The ***combination*** of external representations and physical tools have greatly extended and supported people's ability to carry out cognitive activities (Norman, 1993). Indeed, they are such an integral part that it is difficult to imagine how we would go about much of our everyday life without them.

External cognition is concerned with explaining the cognitive processes involved when we interact with different external representations (Scaife and Rogers, 1996). A main goal is to explicate the cognitive benefits of using different representations for different cognitive activities and the processes involved. The main ones include:

1. externalizing to reduce memory load
2. computational offloading
3. annotating and cognitive tracing

1. Externalizing to reduce memory load

A number of strategies have been developed for transforming knowledge into external representations to reduce memory load. One such strategy is externalizing things we find difficult to remember, such as birthdays, appointments, and addresses. Diaries, personal reminders and calendars are examples of cognitive artifacts that are commonly used for this purpose, acting as external reminders of what we need to do at a given time (e.g., buy a card for a relative's birthday).

Other kinds of external representations that people frequently employ are notes, like "stickies," shopping lists, and to-do lists. Where these are placed in the environment can also be crucial. For example, people often place post-it notes in prominent positions, such as on walls, on the side of computer monitors, by the front door and sometimes even on their hands, in a deliberate attempt to ensure they do remind them of what needs to be done or remembered. People also place things in piles in their offices and by the front door, indicating what needs to be done urgently and what can wait for a while.

Externalizing, therefore, can help reduce people's memory burden by:

- reminding them to do something (e.g., to get something for their mother's birthday)

- reminding them of what to do (e.g., to buy a card)
- reminding them of when to do something (send it by a certain date)

2. Computational offloading

Computational offloading occurs when we use a tool or device in conjunction with an external representation to help us carry out a computation. An example is using pen and paper to solve a math problem.

ACTIVITY 3.6

(a) Multiply 2 by 3 in your head. Easy. Now try multiplying 234 by 456 in your head. Not as easy. Try doing the sum using a pen and paper. Then try again with a calculator. Why is it easier to do the calculation with pen and paper and even easier with a calculator?

(b) Try doing the same two sums using Roman numerals.

Comment

(a) Carrying out the sum using pen and the paper is easier than doing it in your head because you "offload" some of the computation by writing down partial results and using them to continue with the calculation. Doing the same sum with a calculator is even easier, because it requires only eight simple key presses. Even more of the computation has been offloaded onto the tool. You need only follow a simple internalized procedure (key in first number, then the multiplier sign, then next number and finally the equals sign) and then read of the result from the external display.

(b) Using roman numerals to do the same sum is much harder. 2 by 3 becomes II × III, and 234 by 456 becomes CCXXXIIII × CCCCXXXXXVI. The first calculation may be possible to do in your head or on a bit of paper, but the second is incredibly difficult to do in your head or even on a piece of paper (unless you are an expert in using Roman numerals or you "cheat" and transform it into Arabic numerals). Calculators do not have Roman numerals so it would be impossible to do on a calculator.

Hence, it is much harder to perform the calculations using Roman numerals than algebraic numerals—even though the problem is equivalent in both conditions. The reason for this is the two kinds of ***representation*** transform the task into one that is easy and more difficult, respectively. The kind of tool used also can change the nature of the task to being more or less easy.

3. Annotating and cognitive tracing

Another way in which we externalize our cognition is by modifying representations to reflect changes that are taking place that we wish to mark. For example, people often cross things off in a to-do list to show that they have been completed. They may also reorder objects in the environment, say by creating different piles as the nature of the work to be done changes. These two kinds of modification are called annotating and cognitive tracing:

- Annotating involves modifying external representations, such as crossing off or underlining items.

- *Cognitive tracing* involves externally manipulating items into different orders or structures.

Annotating is often used when people go shopping. People usually begin their shopping by planning what they are going to buy. This often involves looking in their cupboards and fridge to see what needs stocking up. However, many people are aware that they won't remember all this in their heads and so often externalize it as a written shopping list. The act of writing may also remind them of other items that they need to buy that they may not have noticed when looking through the cupboards. When they actually go shopping at the store, they may cross off items on the shopping list as they are placed in the shopping basket or cart. This provides them with an annotated externalization, allowing them to see at a glance what items are still left on the list that need to be bought.

Cognitive tracing is useful in situations where the current state of play is in a state of flux and the person is trying to optimize their current position. This typically happens when playing games, such as:

- in a card game, the continued rearrangement of a hand of cards into suits, ascending order, or same numbers to help determine what cards to keep and which to play, as the game progresses and tactics change
- in Scrabble, where shuffling around letters in the tray helps a person work out the best word given the set of letters (Maglio et al., 1999)

It is also a useful strategy for letting users know what they have studied in an online learning package. An interactive diagram can be used to highlight all the nodes visited, exercises completed, and units still to study.

A general cognitive principle for interaction design based on the external cognition approach is to provide external representations at the interface that reduce memory load and facilitate computational offloading. Different kinds of information visualizations can be developed that reduce the amount of effort required to make inferences about a given topic (e.g., financial forecasting, identifying pro-

BOX 3.4 Context-Sensitive Information: Shopping Reminders on the Move

A number of researchers have begun developing wireless communication systems that use GPRS technology to provide people on the move with context-sensitive information. This involves providing people with information (such as reminders and to-do lists) whenever it is appropriate to their location. For example, one such system called comMotion, which is being developed at MIT (Marmasse and Schmandt, 2000) uses a speech-output system to inform people when they are driving past a store that sells the groceries they need to buy, such as milk.

How useful is this kind of externalization? Are people really that bad at remembering things? In what way will it be an improvement over other reminder techniques, like shopping lists written on paper or lists stored on PalmPilots or other pocket computers? Sure, there are certain people who have debilitating memory problems (e.g. Alzheimer's) and who may greatly benefit from having such prosthetic memory devices. But what about those who aren't afflicted? What would happen to them if they started relying more and more on spoken reminders popping up all over the place to tell them what they should be doing when and where? They may well be reminded to buy the milk, but at what price? Losing their own ability to remember?

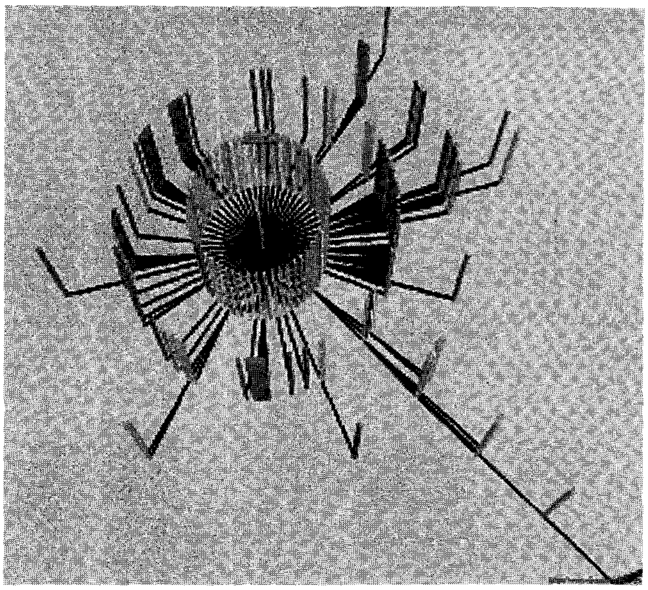

Figure 3.13 Information visualization. Visual Insights' site map showing web page use. Each page appears as a 3D color rod and is positioned radially, with the position showing the location of the page in the site.

gramming bugs). In so doing, they can extend or amplify cognition, allowing people to perceive and do activities that they couldn't do otherwise. For example, a number of information visualizations have been developed that present masses of data in a form that makes it possible to make cross comparisons between dimensions at a glance (see Figure 3.13). GUIs can also be designed to reduce memory load significantly, enabling users to rely more on external representations to guide them through their interactions.

3.5 Informing design: from theory to practice

Theories, models, and conceptual frameworks provide abstractions for thinking about phenomena. In particular, they enable generalizations to be made about cognition across different situations. For example, the concept of mental models provides a means of explaining why and how people interact with interactive products in the way they do across a range of situations. The information processing model has been used to predict the usability of a range of different interfaces.

Theory in its pure form, however, can be difficult to digest. The arcane terminology and jargon used can be quite off-putting to those not familiar with it. It also requires much time to become familiar with it—something that designers and engineers can't afford when working to meet deadlines. Researchers have tried to help out by making theory more accessible and practical. This has included translating it into:

- design principles and concepts
- design rules
- analytic methods
- design and evaluation methods

A main emphasis has been on transforming theoretical knowledge into tools that can be used by designers. For example, Card et al's (1983) psychological model of the human processor, mentioned earlier, was simplified into another model called GOMS (an acronym standing for goals, operators, methods, and selection rules). The four components of the GOMS model describe how a user performs a computer-based task in terms of goals (e.g., save a file) and the selection of methods and operations from memory that are needed to achieve them. This model has also been transformed into the keystroke level method that essentially provides a formula for determining the amount of time each of the methods and operations takes. One of the main attractions of the GOMS approach is that it allows quantitative predictions to be made (see Chapter 14 for more on this).

Another approach has been to produce various kinds of design principles, such as the ones we discussed in Chapter 1. More specific ones have also been proposed for designing multimedia and virtual reality applications (Rogers and Scaife, 1998). Thomas Green (1990) has also proposed a framework of cognitive dimensions. His overarching goal is to develop a set of high-level concepts that are both valuable and easy to use for evaluating the designs of informational artifacts, such as software applications. An example dimension from the framework is "viscosity," which simply refers to resistance to local change. The analogy of stirring a spoon in syrup (high viscosity) versus milk (low viscosity) quickly gives the idea. Having understood the concept in a familiar context, Green then shows how the dimension can be further explored to describe the various aspects of interacting with the information structure of a software application. In a nutshell, the concept is used to examine "how much extra work you have to do if you change your mind." Different kinds of viscosity are described, such as knock-on viscosity, where performing one goal-related action makes necessary the performance of a whole train of extraneous actions. The reason for this is constraint density: the new structure that results from performing the first action violates some constraint that must be rectified by the second action, which in turn leads to a different violation, and so on. An example is editing a document using a word processor without widow control. The action of inserting a sentence at the beginning of the document means that the user must then go through the rest of the document to check that all the headers and bodies of text still lie on the same page.

DILEMMA Evolutionary versus Revolutionary Upgrading

A constant dilemma facing designers involved in upgrading software is where and how to place new functions. Decisions have to be made on how to incorporate them with the existing interface design. Do they try to keep the same structure and add more buttons/menu options, or do they design a new model of interaction that is better suited to organizing and categorizing the increased set of functions? If the former strategy is followed, users do not have to learn a new conceptual model every time they upgrade a piece of software. The down side of trying to keep the same interface structure, however, is that it can easily get overloaded.

A problem when upgrading software, therefore, is working out how to redesign the interaction so that the amount of relearning, relative to the gains from the new functionality, is acceptable by users.

Assignment

The aim of this assignment is for you to elicit mental models from people. In particular, the goal is for you to understand the nature ofpeople's knowledge about an interactive product in terms of how to use it and ***how it works.***

(a) First, elicit your own mental model. Write down how you think a cash machine (ATM) works. Then answer the following questions (abbreviated from Payne, 1991):

How much money are you allowed to take out?

If you took this out and then went to another machine and tried to withdraw the same amount, what would happen?

What is on your card?

How is the information used?

What happens if you enter the wrong number?

Why are there pauses between the steps of a transaction?

How long are they? What happens if you type ahead during the pauses?

What happens to the card in the machine?

Why does it stay inside the machine?

Do you count the money? Why?

Next, ask two other people the same set of questions.

(b) Now analyze your answers. Do you get the same or different explanations? What do the findings indicate? How accurate are people's mental models of the way ATMs work? How transparent are the ATM systems they are talking about?

(c) Next, try to interpret your findings with respect to the design of the system. Are any interface features revealed as being particularly problematic? What design recommendations do these suggest?

(d) Finally, how might you design a better conceptual model that would allow users to develop a better mental model of ATMs (assuming this is a desirable goal)?

This exercise is based on an extensive study carried out by Steve Payne on people's mental models of ATMs. He found that people do have mental models of ATMs, frequently resorting to analogies to explain how they work. Moreover, he found that people's explanations were highly variable and based on ad hoc reasoning.

Summary

This chapter has explained the importance of understanding users, especially their cognitive aspects. It has described relevant findings and theories about how people carry out their everyday activities and how to learn from these when designing interactive products. It has provided illustrations of what happens when you design systems with the user in mind and what happens when you don't. It has also presented a number of conceptual frameworks that allow ideas about cognition to be generalized across different situations.

Key points

- Cognition comprises many processes, including thinking, attention, learning, memory, perception, decision-making, planning, reading, speaking, and listening.