CHAPTER 7

Our Attention is Limited; Our Memory is Imperfect

Using an app, website, or digital appliance requires controlling and focusing our *attention* and using our *memory*.

We focus *attention* on what we are trying to achieve. As we work, the app, website, or appliance produces results—visual, tactile, audible—that (hopefully) capture our attention and allow us to determine whether we have achieved or are approaching our goal.

Our *memory* comes into play in several ways. First, it allows us to *learn* to use the technology: every time we use a digital product or service, we apply what we *remember* from previous experience, reading instructions, or what another person told us, and we store new information. Our memory allows us to retain the *knowledge* of how to use the technology from one session to the next, and it allows us to increase our knowledge over time. In addition, our memory—augmented by feedback from the app, website, or appliance—allows us to keep track of our goal and what we have already done.

Because memory and attention are crucial components in people's use of digital technology, it is important for designers of digital products and services to know something about them. That is what this chapter is about.

Just as the human visual system has strengths and weaknesses, so do human attention and memory. This chapter describes some of those strengths and weaknesses as the background for understanding how to design interactive systems to support and augment attention and memory rather than burdening or confusing them. We will start with an overview of how memory works and how it is related to attention.

SHORT- VERSUS LONG-TERM MEMORY

Psychologists historically have distinguished *short-term* memory from *long-term* memory. Short-term memory covers situations in which information is retained for intervals ranging from a fraction of a second to a few minutes. Long-term memory

Designing with the Mind in Mind. https://doi.org/10.1016/B978-0-12-818202-4.00007-6

FIGURE 7.1

Traditional (antiquated) view of short-term versus long-term memory.

covers situations in which information is retained over longer periods (e.g., hours, days, years, or even lifetimes).

It is tempting to think of short- and long-term memory as separate memory stores. Indeed, some theories of memory have considered them separate. After all, in a digital computer, the short-term memory stores (central processing unit data registers) are separate from the long-term memory stores (random-access memory, hard disk, flash memory, CD-ROM, etc.). More direct evidence comes from findings that damage to certain parts of the human brain results in short-term but not long-term memory deficits, or vice versa. Finally, the speed with which information or plans can disappear from our immediate awareness contrasts sharply with the seeming permanence of our memory of important events in our lives, faces of significant people, activities we have practiced, and information we have studied. These phenomena led many researchers to theorize that short-term memory is a separate store in the brain where information is held temporarily after entering through our perceptual senses (e.g., visual or auditory) or after being retrieved from long-term memory (see Fig. 7.1).

A MODERN VIEW OF MEMORY

Recent research on memory and brain function indicates that short- and long-term memory are functions of a single memory system—one more closely linked with perception than previously thought (Jonides et al., 2008).

Long-term memory

Perceptions enter through the visual, auditory, olfactory, gustatory, or tactile sensory systems and trigger responses starting in areas of the brain dedicated to each sense (e.g., visual cortex, auditory cortex), then spread into other areas of the brain *not* specific to any particular sensory modality. The sensory modality-specific areas of the brain detect only simple features of the data, such as a dark/light edge, diagonal line, high-pitched tone, sour taste, red color, or rightward motion. Downstream areas of the brain combine low-level features to detect higher-level features of the input, such as animal, the word "duck," Uncle Kevin, minor key, threat, or fairness.

As described in Chapter 1, the set of neurons activated by a perceived stimulus depends on both the *features* and the *context* of the stimulus. Context is as important as features in determining what neural patterns are activated. For example, a dog barking near you when you are walking in your neighborhood activates a different pattern of neural activity in your brain than the same sound heard when you are listening to a recording. The more similar that two perceptual stimuli are—that is, the more features and contextual elements they share—the more overlap there is between the sets of neurons that fire in response to them.

The initial strength of a perception depends on how much it is amplified or dampened by other brain activity. All perceptions create some kind of trace, but some are so weak that they can be considered not registered: the pattern was activated once but never again.

Memory formation consists of changes in the neurons involved in a neural activity pattern that make the pattern easier to reactivate in the future.[1] Some such changes result from chemicals released near neural endings that boost or inhibit their sensitivity to stimulation. These changes last only until the chemicals dissipate or are neutralized by other chemicals. More permanent changes occur when neurons grow and branch, forming new connections with others.

Activating a memory consists of reactivating the same pattern of neural activity that occurred when the memory was formed. Somehow the brain distinguishes initial activations of neural patterns from *re*activations—perhaps by measuring the relative ease with which the pattern was reactivated. New perceptions very similar to the original ones reactivate the same patterns of neurons, resulting in *recognition* if the reactivated perception reaches awareness. In the absence of a similar perception, stimulation from activity in other parts of the brain can also reactivate a pattern of neural activity, which if it reaches awareness results in *recall*.

The more often a neural memory pattern is reactivated, the stronger it becomes—that is, the easier it is to reactivate—which in turn means that the perception it corresponds to is easier to recognize and recall. Neural memory patterns can also be strengthened or weakened by excitatory or inhibitory signals from other parts of the brain.

A particular memory is not located in any specific spot in the brain. The neural activity pattern comprised in a memory involves a network of millions of neurons extending over a wide area. Activity patterns for different memories overlap depending on which features they share. Removing, damaging, or inhibiting neurons in a particular part of the brain typically does not completely wipe out memories that involve those neurons but rather just reduces the detail or accuracy of the memory by deleting features.[2] However, some areas in a neural activity pattern may be critical pathways, so removing, damaging, or inhibiting them may prevent most of the pattern from activating, thereby effectively eliminating the corresponding memory.

---

[1] There is evidence that the long-term neural changes associated with learning occur mainly during sleep, suggesting that separating learning experiences by periods of sleep facilitates learning (Stafford and Webb, 2005).

[2] This is similar to the effect of cutting pieces out of a holographic image: it reduces the overall resolution of the image, rather than removing areas of it as with an ordinary photograph.

For example, researchers have long known that the hippocampus, twin seahorse-shaped neural clusters near the base of the brain, plays an important role in storing long-term memories. The modern view is that the hippocampus is a controlling mechanism that directs neural rewiring to "burn" memories into the brain's wiring. The amygdala, two jellybean-shaped clusters on the frontal tips of the hippocampus, has a similar role, but it specializes in storing memories of emotionally intense, threatening situations (Eagleman, 2012).

Cognitive psychologists view human long-term memory as consisting of three distinct functions:

- ***Semantic*** long-term memory stores *facts and relationships*, such as "Germany is in Europe," "Berlin is the capital of Germany," "dogs are mammals," and "2 + 2 = 4."
- ***Episodic*** long-term memory records *past events*, such as the time you hit a baseball through a neighbor's window, the time you met your spouse, or yesterday's dinner.
- ***Procedural*** long-term memory remembers *action sequences*, such as how to tie your shoes, execute a karate move, or play a song on your guitar.

Everything we do, including using digital technology, uses all three types of long-term memory. For example, when you text a friend on your smartphone, you use semantic memory to generate what you want to say, episodic memory to recall what you and your friends did yesterday, and procedural memory to guide your hands in tapping out the text of your message and sending it. Furthermore, your phone's texting app *supports* all three types of memory. It supports semantic memory by helping you keep track of your friends and their likes and dislikes and by sending you facts to add to your knowledge; it supports episodic memory by showing you texts you have sent and received as well as by providing news about events, and it supports procedural memory—or perhaps hampers it—by autocompleting many of the words you type.

Short-term memory

So far, we've discussed long-term memory. What about short-term memory? What psychologists call short-term memory is actually a *combination* of phenomena involving perception, attention, and retrieval from long-term memory.

One component of short-term memory is perceptual. Each of our perceptual senses has its own very brief short-term "memory" resulting from residual neural activity after a perceptual stimulus ceases, like a bell that rings briefly after it is struck. Until they fade away, these residual perceptions are available as possible input to our brain's attention and memory-storage mechanisms, which integrate input from our various perceptual systems, focus our awareness on some of that input, and store some of it in long-term memory. These sensory-specific residual perceptions together make up a minor component of short-term memory. Here, we are only interested in them as potential inputs to working memory.

Also available as potential input to working memory are long-term memories reactivated through recognition or recall. As explained earlier, each long-term memory corresponds to a specific pattern of neural activity distributed across our brain. While activated, a memory pattern is a candidate for our attention and therefore potential input for working memory.

The human brain has multiple attention mechanisms, some voluntary and some involuntary. They focus our awareness on a very small subset of perceptions and activated long-term memories while ignoring everything else. The tiny subset of all available information from our perceptual systems and long-term memories that we are aware of *right now* is the main component of our short-term memory, the part that cognitive scientists often call *working memory*. It integrates information from all our sensory modalities and long-term memory. Henceforth, we will restrict our discussion of short-term memory to working memory.

So what is working memory? First, here is what it is *not*: it is not a store—it is not a *place* in the brain where memories and perceptions *go* to be worked on. And it is nothing like accumulators or fast random-access memory in digital computers.

Instead, working memory is the combined focus of our attention: everything that we are conscious of at a given time. More precisely, it is a few perceptions and long-term memories that are activated enough that we remain aware of them over a short period. Psychologists also view working memory as including an executive function—based mainly in the frontal cerebral cortex—that manipulates items we are attending to, and if needed refreshes their activation so they remain in our awareness (Baddeley, 2012).

A useful—if oversimplified—analogy for memory is a huge, dark, musty warehouse. The warehouse is full of long-term memories piled haphazardly (not stacked neatly), intermingled and tangled, and mostly covered with dust and cobwebs. Doors along the walls represent our perceptual senses: sight, hearing, smell, taste, touch. They open briefly to let perceptions in. As perceptions enter, they are briefly illuminated by light coming in from outside, but they quickly are pushed (by more entering perceptions) into the dark tangled piles of old memories.

In the ceiling of the warehouse are a small fixed number of searchlights controlled by the attention mechanism's executive function (Baddeley, 2012). They swing around and focus on items in the memory piles, illuminating them for a while until they swing away to focus elsewhere. Sometimes one or two searchlights focus on new items after they enter through the doors. When a searchlight moves to focus on something new, whatever it had been focusing on is plunged into darkness.

The small fixed number of searchlights represents the limited capacity of working memory. What is illuminated by them (and briefly through the open doors) represents the contents of working memory—out of the vast warehouse's entire contents, the few items we are attending to at any moment (see Fig. 7.2).

The warehouse analogy is too simple and should not be taken too seriously. As Chapter 1 explained, our senses are *not* just passive doorways into our brains through which our environment "pushes" perceptions. Rather, our brain actively and continually *seeks out* important events and features in our environment and

FIGURE 7.2

Modern view of memory: a dark warehouse full of stuff (long-term memory) with searchlights focused on a few items (short-term memory).

"pulls" perceptions in as needed (Ware, 2008). Furthermore, the brain is buzzing with activity most of the time, and its internal activity is only modulated—not determined—by sensory input (Eagleman, 2012). Also, as described earlier, memories are embodied as networks of neurons distributed around the brain, not as objects in a specific location. Finally, activating a memory in the brain can activate related ones; our warehouse-with-searchlights analogy doesn't represent that.

Nonetheless, the analogy—especially the part about the searchlights—illustrates that working memory is a *combination* of several *foci of* attention (the currently activated neural patterns of which we are aware), its capacity is extremely limited, and its content at any given moment is very volatile.

What about the earlier finding that damage to some parts of the brain causes short-term memory deficits, while other types of brain damage cause long-term memory deficits? The current interpretation is that some types of damage decrease or eliminate the brain's ability to focus attention on specific objects and events, while other types of damage harm the brain's ability to store or retrieve long-term memories.

CHARACTERISTICS OF ATTENTION AND WORKING MEMORY

As noted, working memory is equal to the focus of our attention. Whatever is in that focus is what we are conscious of at any moment. But what determines what we attend to and how much we can attend to at any given time?

Attention is highly focused and selective

Most of what is going on around you at this moment you are unaware of. Your perceptual system and brain sample very selectively from your surroundings because they don't have the capacity to process everything.

Right now you are conscious of the last few words and ideas you have read but probably not the color of the wall in front of you. But now that I have shifted your attention, you *are* conscious of the wall's color and may have forgotten some of the ideas you read on the previous page.

Chapter 1 described how our perception is filtered and biased by our goals. For example, if you are looking for your friend in a crowded shopping mall, your visual system "primes" itself to notice people who look like your friend (including how he or she is dressed), and barely notice everything else. Simultaneously, your auditory system primes itself to notice voices that sound like your friend's voice and even footsteps that sound like those of your friend. Human-shaped blobs in your peripheral vision and sounds localized by your auditory system that match those of your friend snap your eyes and head toward them. While you look, anyone looking or sounding similar to your friend attracts your attention, and you won't notice other people or events that would normally have interested you.

Besides focusing on objects and events related to our current goals, our attention is drawn to:

- ***Movement, especially movement near or toward us***. For example, something jumps at you while you walk on a street, or something swings toward your head in a haunted house ride at an amusement park, or a car in an adjacent lane suddenly swerves toward your lane (see the discussion of the *flinch reflex* in Chapter 14).
- ***Threats***. Anything that signals or portends danger to us or people in our care.
- ***Faces of other people***. We are primed from birth to notice faces more than other objects in our environment.
- ***Sex and food***. Even if we are happily married and well fed, these things attract our attention. Even the mere words probably got your attention.

These things, along with our current goals, draw our attention involuntarily. We don't become aware of something in our environment and then orient ourselves toward it. It is the other way around: our perceptual system detects something attention-worthy and orients us toward it preconsciously, and only afterward do we become aware of it.[3]

---

[3] Exactly *how long* afterward is discussed in Chapter 14.

Working memory (aka attention) is limited

The primary characteristics of working memory are its low capacity and volatility. But what is that capacity? In terms of the warehouse analogy presented earlier, what is *the small fixed number* of searchlights?

Many college-educated people have read about "the magical number seven, plus or minus two" proposed by cognitive psychologist George Miller in 1956 as the limit on the number of simultaneous unrelated items in human working memory (Miller, 1956).

Miller's characterization of the working-memory limit naturally raises several questions:

- ***What are the items in working memory?*** They are current perceptions and retrieved memories. They are goals, numbers, words, names, sounds, images, odors—anything one can be aware of. In the brain, they are patterns of neural activity.
- ***Why must items be unrelated?*** Because if two items are related, they correspond to one big neural activity pattern—one set of features—and hence one item, not two.
- ***Why the fudge factor of plus or minus two?*** Because researchers cannot measure with perfect accuracy how much people can keep track of as well as individual differences in working-memory capacity.

Later research in the 1960s and 1970s found Miller's estimate to be too high. In the experiments Miller considered, some of the items presented to people to remember could be "chunked" (i.e., considered related), making it appear that people's working memories were holding more items than they actually were. Furthermore, all the subjects in Miller's experiments were college students. Working-memory capacity varies within the general population. When the experiments were revised to disallow unintended chunking and include noncollege students as subjects, the average capacity of working memory was shown to be more like four plus or minus one—that is, three to five items (Broadbent, 1975; Mastin, 2010). Thus, in our warehouse analogy, there would be only four searchlights.

More recent research has cast doubt on the idea that the capacity of working memory should be measured in whole items or "chunks." It turns out that in early working-memory experiments, people were asked to briefly remember items (e.g., words or images) that were quite different from each other—that is, they had very few features in common. In such a situation, people don't have to remember every feature of an item to recall it a few seconds later; remembering some of its features is enough. So people appeared to recall items as a whole, and therefore working-memory capacity seemed measurable in whole items.

Recent experiments have given people items to remember that are similar—that is, they share many features. In that situation, to recall an item and not confuse it with other items, people must remember more of its features. In these experiments, researchers found that people remember more details (i.e., features) of some items than of others, and the items they remember in greater detail are the ones they paid more attention to (Bays and Husain, 2008). This suggests that the unit of attention—and therefore the capacity of working memory—is best measured in item features rather than whole items or "chunks" (Cowan et al., 2004). This jibes with the modern view of the brain as a feature-recognition device, but it is controversial among memory researchers, some of whom argue that the basic capacity of human working memory is three to five whole items but is reduced when people attend to a large number of the details (i.e., features) of items (Alvarez and Cavanagh, 2004).

Bottom line: The true capacity of human working memory is still a research topic.

The second important characteristic of working memory is how volatile it is. Cognitive psychologists used to say that new items arriving in working memory often bump old ones out, but that way of describing the volatility is based on the view of working memory as a temporary storage place for information. The modern view of working memory as the current focus of attention makes it even clearer: focusing attention on new information turns it away from some of what it was focusing on. That is why the searchlight analogy is useful.

However we describe it, information can easily be lost from working memory. If items in working memory don't get combined or rehearsed, they are at risk of having the focus shifted away from them. This volatility applies to goals as well as to the details of objects. Losing items from working memory corresponds to forgetting or losing track of something you were doing. We have all had such experiences, for example:

- Going to another room for something, but once there we cannot remember why we came.
- Taking a phone call and afterward not remembering what we were doing before the call.
- Something yanks our attention away from a conversation and then we can't remember what we were talking about.
- In the middle of adding a long list of numbers, something distracts us, so we have to start over.

WORKING MEMORY TEST

To test your working memory, copy these instructions onto a separate sheet of paper, get a pen or pencil and two blank sheets of paper, and follow these instructions:

1. Flip to the next page of the book but don't look at it. Cover it with a blank sheet of paper.
2. Pull the paper cover down a little, read the **black numbers** at the top for a few seconds, then slide the blank sheet back up to cover the book page again. *Don't peek ahead at other numbers on that page unless you want to ruin the test.* Look back at these instructions.
3. Say your phone number backward, out loud.
4. Now write down the black numbers from memory … now check the black numbers again. Did you get all of them?
5. Pull the cover down a bit more, read the **red numbers** (under the black ones), and look back at these instructions.
6. Write down the numbers from memory. These would be easier to recall than the first ones if you noticed that they are the first seven digits of π (3.141,592), because then they would be only one number, not seven.
7. Return to the book page, read the **green numbers** for 3 seconds, then look back here.
8. Write down the numbers you remember. If you noticed that they are odd numbers from 1 to 13, they will be easier to recall because you only need to remember three chunks ("odd #s, 1 to 13" or "seven odd #s, from 1"), not seven.
9. Return to the book page, read the **orange words** for 3 seconds, then look back here.
10. Write down the words from memory … could you recall them all?
11. Return to the book page, read the **blue words** for 3 seconds, then look back here.
12. Write down the words from memory … it was a lot easier to recall them because they form a sentence, so they could be stored as one sentence rather than seven words.

3 8 4 7 5 3 9

3 1 4 1 5 9 2

1 3 5 7 9 11 13

town river corn string car shovel

what is the meaning of life

IMPLICATIONS OF WORKING-MEMORY CHARACTERISTICS FOR USER-INTERFACE DESIGN

The capacity and volatility of working memory have many implications for the design of interactive computer systems. The basic implication is that user interfaces should help people remember essential information from one moment to the next. Don't require people to remember system status or what they have done, because their attention is focused on their primary goal and progress toward it. Specific examples follow.

Voice user interfaces

Voice user interfaces (VUIs) place a higher load on short-term memory than graphical user interfaces do (Pearl, 2018a). The most obvious design implication of the limited capacity of human working memory is that computer systems that present information via speech output should limit how much information they provide at once so they don't overload users' short-term memory. This includes telephone voice-response systems like those that companies provide to help you before they let you speak with a real person. It also includes personal digital assistants like Siri, Alexa, Google Assistant, and Cortana.

For example, VUIs should *not* issue long instructions in a single long utterance. Bad: *"Welcome to ACME travel service, your gateway to anywhere. I can help you with flight reservations, hotel bookings, or car rentals. Just tell me which one* *you want. If you need help, say* 'help'. *Unfortunately, our car rental service is not available now. What would you like?"* Few people will be able to remember all of that. Instead, get rid of nonessential text and break the interaction up into short segments. Better: *"ACME travel. Say* 'flight reservation'*,* 'book hotel'*, or* 'rent car'*."* (User says "rent car.") "OK, you want to rent a car. In what city?" (etc.)

A related VUI design guideline is that verbal lists of options or search results should not be longer than four or five items. People can't remember more than that (Budiu and Laubheimer, 2018) unless the items are familiar, such as close friends, family members, or favorite restaurants (Pearl, 2018b).

Some VUIs run on devices that also have screens, such as smartphones, tablets, laptop computers, and some smart speakers. In such cases, the VUI can supplement its verbal output by displaying information on the screen, thus reducing what it needs to say and thereby decreasing the burden on users' short-term memory (Pearl, 2018a).

Modes

The limited capacity and volatility of working memory is one reason user-interface design guidelines often say to either avoid designs that have *modes* or provide adequate mode feedback. In a moded user interface, some user actions have different effects depending on what mode the system is in. For example:

- In a car, pressing the accelerator pedal can move the car either forward, backward, or not at all depending on whether the transmission is in drive, reverse, or neutral. The transmission sets a mode in the car's user interface.
- In many digital cameras, pressing the shutter button can either snap a photo or start a video recording depending on which mode is selected.
- In a drawing program, clicking and dragging normally selects one or more graphic objects on the drawing, but when the software is in "draw rectangle" mode, clicking and dragging adds a rectangle to the drawing and stretches it to the desired size.

Moded user interfaces have advantages; that is why many interactive systems have them. Modes allow a device to have more functions than controls: the same control provides different functions in different modes. Modes allow an interactive system to assign different meanings to the same gestures to reduce the number of gestures users must learn.

However, one well-known *dis*advantage of modes is that people often make *mode errors*: they forget what mode the system is in and do the wrong thing by mistake (Johnson, 1990). This is especially true in systems that give poor feedback about what mode the system is in. Because of the problem of mode errors, many user-interface design guidelines say to either avoid modes or provide strong feedback about which mode is the current one. Working memory is too unreliable for designers to assume that users can, without clear, continuous feedback, keep track of what mode the system is in, even when the users are the ones changing modes.

Search results

When people use a search function on a computer to find information, they enter the search terms, start the search, and then review the results. Evaluating the results often requires knowing what the search terms were. If working memory were less limited, people would always remember, when browsing the results, what they had entered as search terms just a few seconds earlier. But as we have seen, working memory is very limited. When the results appear, a person's attention naturally turns away from what he or she entered and toward the results. Therefore, it should be no surprise that people viewing search results often do not remember the search terms they just typed.

Unfortunately, some designers of online search functions don't understand that. Search results sometimes do not show the search terms that generated the results. For example, in 2006, the search results page at Slate.com provided search fields so users could search again, but didn't show what a user had searched for (see Fig. 7.3A). The 2020 version of the site shows the user's search terms (see Fig. 7.3B), reducing the burden on user working memory.

Calls to action

A well-known "netiquette" guideline for writing email messages, especially messages that require responses or ask the recipients to do something, is to restrict each message to one topic. If a message contains multiple topics or requests, its recipients may focus on one of them (usually the first one), get engrossed in responding to that, and forget to respond to the rest of the email. The guideline to put different topics or requests into separate emails is a direct result of the limited capacity of human attention.

Web designers are familiar with a similar guideline: avoid putting competing calls to action on a page. Each page should have only *one* dominant call to action—or one for each possible user goal—to not overwhelm users' attention capacity and cause them go down paths that don't achieve their (or the site owner's) goals. A related guideline: once users have specified their goal, don't distract them from accomplishing it by displaying extraneous links and calls to action. Instead, guide them to the goal by using a design pattern called the *process funnel* (van Duyne et al., 2002; see also Johnson, 2007).

Navigation depth

Using a software product, digital device, phone menu system, or website often involves navigating to the user's desired information or goal. It is well established that broad, shallow navigation hierarchies are easier for most people—especially those who are nontechnical—to find their way around in than narrow, deep hierarchies (Cooper, 1999). This applies to hierarchies of application windows and dialogue boxes as well as to menu hierarchies (Johnson, 2007).

A related guideline: in hierarchies deeper than two levels, provide navigation "bread crumb" paths to constantly remind users where they are (Nielsen, 1999; van Duyne et al., 2002).

These guidelines, like those mentioned earlier, are based on the limited capacity of human working memory. Requiring a user to drill down through eight levels of dialogue boxes, web pages, menus, or tables—especially with no visible reminders of their location—will probably exceed the user's working-memory capacity, thereby causing him or her to forget where he or she came from or what his or her overall goals were.

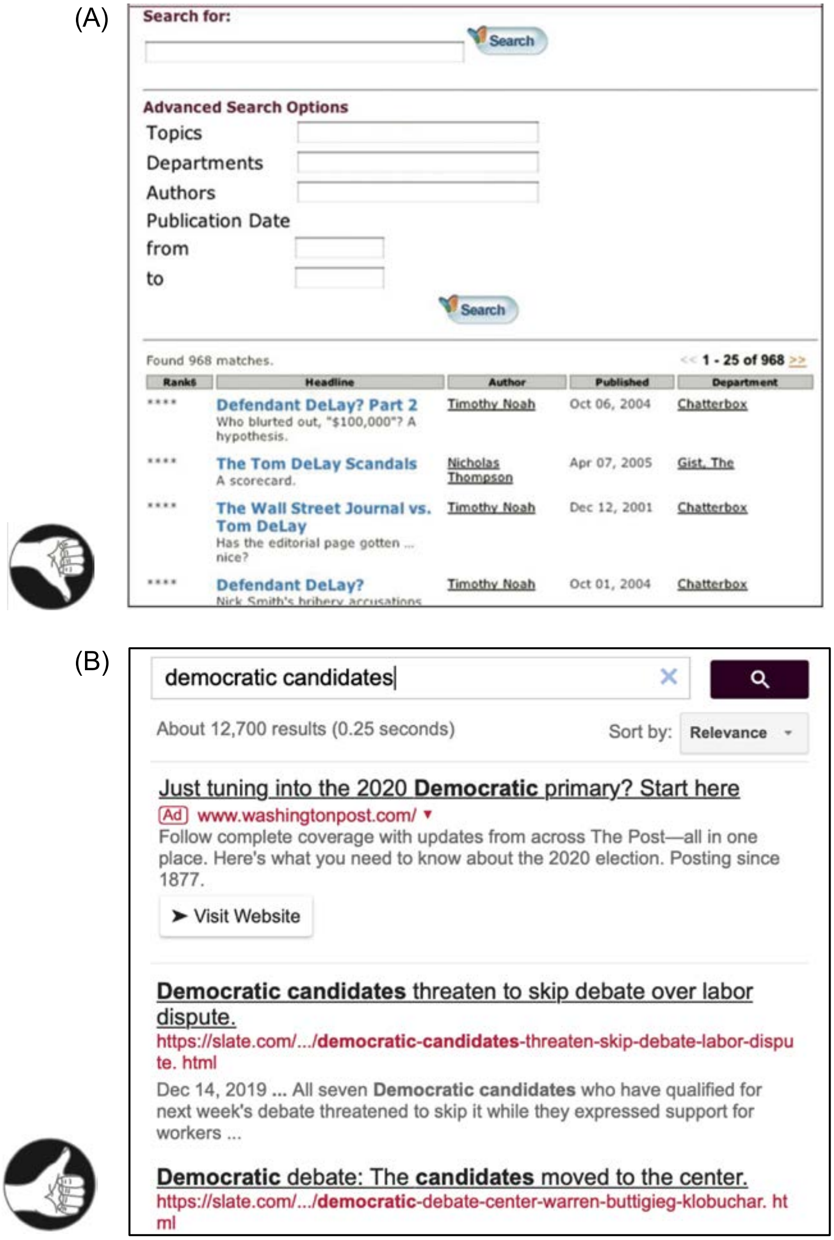

FIGURE 7.3

Slate.com search results: in 2007 (A), users' search terms were not shown, but in 2020 (B), search terms are shown.

CHARACTERISTICS OF LONG-TERM MEMORY

Long-term memory differs from working memory in many respects. Unlike working memory, it actually *is* a memory store.

However, specific memories are not stored in any one neuron or location in the brain. As described earlier, memories, like perceptions, consist of activation patterns of large sets of neurons, with subsets of neurons and individual neurons in each pattern encoding specific features of the memory. Related memories correspond to overlapping patterns of activated neurons, because similar memories share details—i.e., features. For example, my tablet computer is different from my smartphone, but my memories of how to read and send email on the two devices share features because the devices and the email apps on them are similar, so the neural firing patterns that encode both memories overlap quite a lot. Every memory is stored in a distributed fashion, spread among many parts of the brain. In this way, long-term memory in the brain is similar to holographic light images.

Long-term memory evolved to serve our ancestors and us very well in getting around in our world. However, it has many weaknesses: it is error-prone, impressionist, free-associative, idiosyncratic, retroactively alterable, and easily biased by a variety of factors at the time of recording or retrieval. Let's examine some of these weaknesses.

We will begin our examination of long-term memory weaknesses by considering the capacity of human long-term memory. Unlike working memory, the capacity of human long-term memory seems almost unlimited. No human has ever run out of "memory space." Why?

Before we answer that question, let's estimate how much information a human brain can hold. Adult human brains each contain about 86 billion neurons (Herculano-Houzel, 2009). If each neuron encodes 1 bit of memory, that suggests that the information capacity of the human brain is about 86 gigabits, which is 10.75 gigabytes. By that reckoning, the human brain has less memory than most mobile phones.

But wait: neurons can vary their firing rate, so maybe a neuron can represent several bits, which suggests a *larger* total information capacity. On the other hand, not all neurons in the brain are involved in storing memories, suggesting a *lower* total capacity. However, as described earlier, individual neurons do not store memories; memories are encoded by *networks* of neurons acting together. Even if only *some* of the brain's neurons are involved in memory, the large number of those neurons and the fact that each one can participate in many neural networks allows for an astronomical number of combinations of neurons into networks, with each network potentially encoding a different memory. By that reckoning, the maximum memory-capacity of the human brain is much, much greater than what we estimated above—perhaps terabytes. The overall result is that researchers have not yet measured or even provided a plausible estimate of the maximum information capacity of the human brain.[4]

[4] The closest researchers have come is Landauer's (1986) use of the average human learning rate to calculate the amount of information a person can learn in the average lifetime: 10^9 bits, or a few hundred megabytes; not much compared with today's computers, phones, and flash drives.

However, for practical purposes it doesn't matter what the maximum information capacity of the human brain may be at any instant in time, because the brain does not store information like a digital computer. New memories "steal" neurons from old memories (Eagleman, 2015). A new experience is never totally new; it shares features with prior experiences, so the network it activates shares neurons with the networks that encode those older experiences. Over time, a new memory is strengthened by repeated experience or practice, while related older memories lose details—researchers say that old memories lose *resolution*.

For example, if you don't see someone you know for months or years, your memory of their face slowly loses details—features—as your brain reuses some of the neurons that encode that memory to encode new face-memories. Similarly, as you learn to use a new computer or smartphone app, you will forget details of how to use related apps if you don't practice using them too. To paraphrase an old saying: "Old memories don't disappear; they just lose details."

Bottom line: nearly everything we've ever experienced is stored in our long-term memory but at different degrees of detail, and the detail of each memory can increase with practice and decrease with disuse.

Error-prone

Long-term memory is not an accurate, high-resolution recording of our experiences. In terms familiar to computer engineers, one could characterize long-term memory as using heavy compression methods that drop a great deal of information. Images, concepts, events, sensations, actions—all are reduced to combinations of features. Different memories are stored at different levels of detail—that is, with more or fewer features. Also, memories in the human brain can be changed by the retrieval process; that is, *reading* a memory also *writes* into it. That is very unlike memory in digital computers, in which reading and writing are distinct operations.

For example, the face of a man you met briefly who is not important to you might be stored simply as an average Caucasian male face with a beard with no other details—a whole face reduced to three features. If you were asked later to describe the man in his absence, the most you could honestly say was that he was a "white guy with a beard." You would not be able to pick him out of a police lineup of other Caucasian men with beards. In contrast, your memory of your best friend's face includes many more features, allowing you to give a more detailed description and pick your friend out of any police lineup. Nonetheless, it is still a set of features, not anything like a bitmap image.

As another example, I have a vivid childhood memory of being run over by a plow and badly cut, but my father claimed that it happened to my brother. One of us is wrong.

In the realm of human–computer interaction, a Microsoft Word user may remember that there is a command to insert a page number but may not remember which *menu* the command is in. That specific feature may not have been recorded when the user learned how to insert page numbers. Alternatively, perhaps the menu-location feature *was* recorded but just does not reactivate with the rest of the memory pattern when the user tries to recall how to insert a page number.

Weighted by emotions

Chapter 1 described a dog that remembered seeing a cat in his front yard every time he returned home in the family car. The dog was excited when he first saw the cat, so his memory of it was strong and vivid.

A comparable human example would be that an adult could easily have strong memories of her first day at nursery school but probably not of her tenth. On the first day, she was probably upset about being left at the school by her parents, whereas by the 10th day, being left there was nothing unusual.

Retroactively alterable

Suppose that while you are on an ocean cruise with your family, you see a whale shark. Years later, when you and your family are discussing the trip, you might remember seeing a whale, and one of your relatives might recall seeing a shark. For both of you, some details in long-term memory were dropped because they did not fit a common concept.

A true example comes from 1983, when the late President Ronald Reagan was speaking with Jewish leaders during his first term as president. He spoke about being in Europe during World War II and helping to liberate Jews from the Nazi concentration camps. The trouble was, he was never in Europe during World War II. When he was an actor, he was in a *movie* about World War II, made entirely in Hollywood. That important detail had been dropped from his memory.

A LONG-TERM MEMORY TEST

Test your long-term memory by answering the following questions:

1. Was there a roll of tape in the toolbox in Chapter 1?
2. What was your *previous* phone number?
3. Which of these words were *not* in the list presented in the working-memory test earlier in this chapter: city, stream, corn, auto, twine, spade?
4. What was your first-grade teacher's name? Second grade? Third grade?
5. What website was presented earlier that does not show search terms when it displays search results?

Regarding question 3: When words are memorized, often what is retained is the *concept* rather than the exact word presented. For example, one could hear the word "town" and later recall it as "city."

IMPLICATIONS OF LONG-TERM MEMORY CHARACTERISTICS FOR USER-INTERFACE DESIGN

The main thing that the characteristics of long-term memory imply is that people need tools to augment it. Since prehistoric times, people have invented technologies to help them remember things over long periods: notched sticks, knotted ropes, mnemonics, verbal stories and histories retold around campfires, writing, scrolls, books, number systems, shopping lists, checklists, phone directories, datebooks, accounting ledgers, oven timers, computers, portable digital assistants (PDAs), online shared calendars, etc.

Given that humankind has a need for technologies that *augment* memory, it seems clear that software designers should try to provide software that fulfills that need. At the very least, designers should avoid developing systems that *burden* long-term memory. Yet that is exactly what many interactive systems do.

Authentication is one functional area in which many software systems place burdensome demands on users' long-term memory. For example, a web application developed a few years ago told users to change their personal identification number (PIN) "to a number that is easy to remember" but then imposed restrictions that made it impossible to do so (see Fig. 7.4). Whoever wrote those instructions seems to have realized that the PIN requirements were unreasonable, because the instructions end by advising users to write down their PIN! Never mind that writing a PIN down creates a security risk and adds yet *another* memory task: users must remember where they hid their written-down PIN.

A contrasting example of burdening people's long-term memory for the sake of security comes from United.com. To make flight reservations, visitors must register. The site requires users to select a security question from a menu (see Fig. 7.5). What if you can't answer *any* of the questions in a unique way that you will recall when the site challenges you? What if you don't recall your childhood favorite fruit or vegetable,

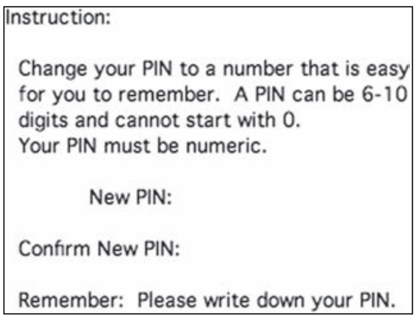

FIGURE 7.4

Instructions tell users to create an easy-to-remember PIN, but the restrictions make that impossible.

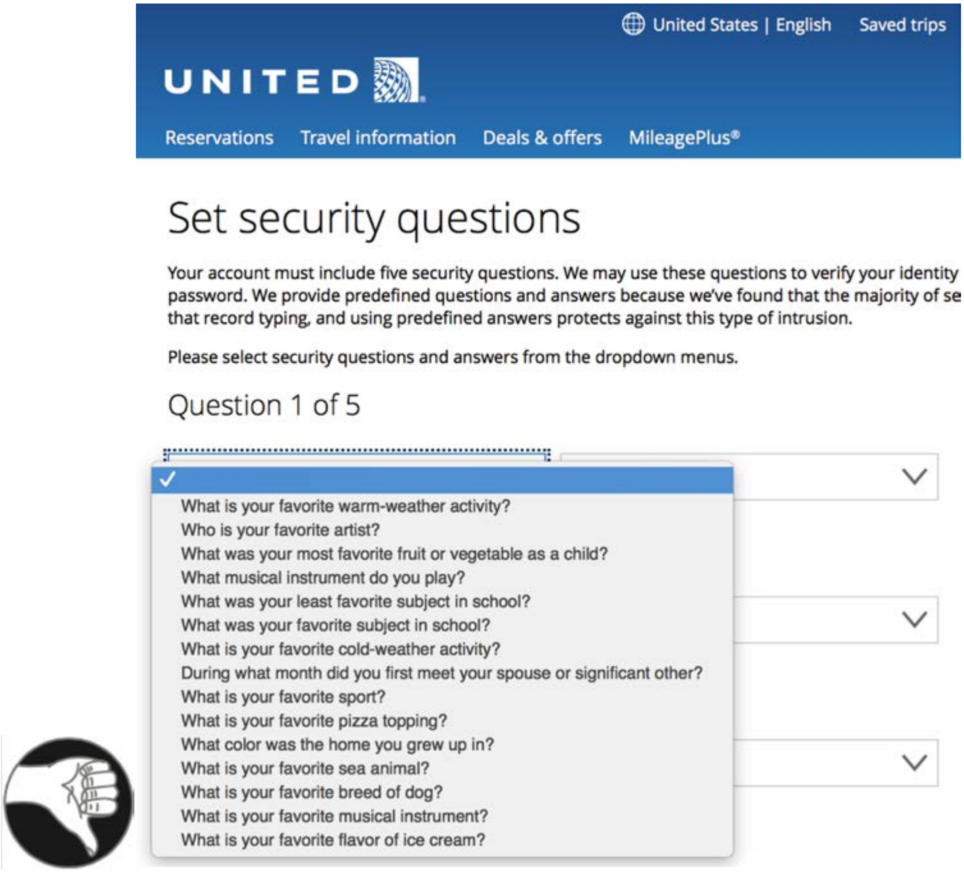

FIGURE 7.5

United.com's registration (2019) burdens long-term memory: a user may not have a unique, memorable answer for any security question.

your favorite subject in school, the color of your childhood home, or the answers to the other questions?

But that isn't where the memory burden ends. Some questions could have several possible answers. Many people had several favorite subjects, childhood homes, or cold-weather activities. To register, they must choose a question and then *remember* which answer they gave to United.com. How? Probably by writing it down somewhere. Then, when United.com asks them the security question, they have to *remember* where they put the answer. Why burden people's memory, when it would be easy to let users make up a security question for which they can easily recall the one possible answer?

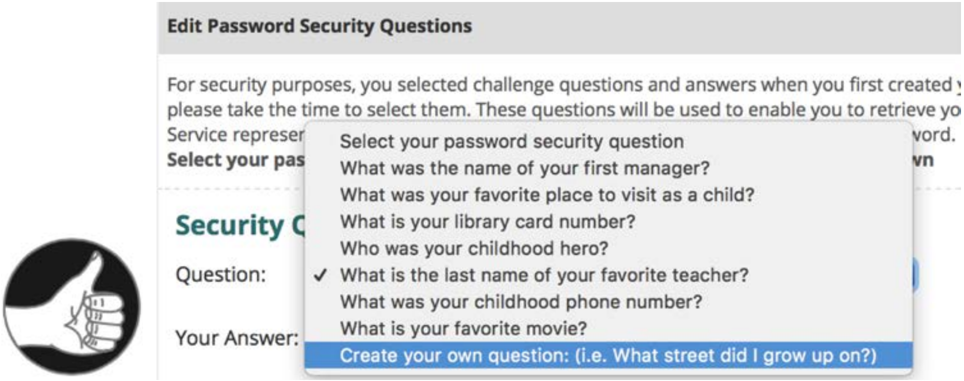

FIGURE 7.6

NetworkSolutions.com lets users create a security question if none on the menu work for them.

Such unreasonable demands on people's long-term memory counteract the security and productivity that computer-based applications supposedly provide (Schrage, 2005), as users:

- place sticky notes on or near computers or "hide" them in desk drawers;
- contact customer support to recover passwords they cannot recall;
- use passwords that are easy for others to guess;
- set up systems with no login requirements at all or with one shared login and password.

The registration form at NetworkSolutions.com represents a small step toward more useable security. Like United.com, it offers a choice of security questions, but it also allows users to create their own security question—one for which they can more easily remember the answer (see Fig. 7.6).

Another implication of long-term memory characteristics for interactive systems is that learning and long-term retention are enhanced by user-interface consistency.

The more consistent the operation of different functions, or the more consistent the actions on different types of objects, the less users have to learn. User interfaces that have many exceptions and little consistency from one function or object to another require users to store in their long-term memory many features about each function or object and its correct usage context. The need to encode more features makes such user interfaces harder to learn. It also makes it more likely that a user's memory will drop essential features during storage or retrieval, increasing the chances that the user will fail to remember, misremember, or make other memory errors.

Even though some have criticized the concept of consistency as ill defined and easy to apply badly (Grudin, 1989), the fact is that consistency in a user interface greatly reduces the burden on users' long-term memory. Mark Twain once wrote, "If you tell the truth, you never have to remember anything." One could also say, "If everything worked the same way, you would not have to remember much." We will return to the issue of user-interface consistency in Chapter 11.

IMPORTANT TAKEAWAYS

- Human attention and memory have strengths and weaknesses. By understanding those strengths and weaknesses, we can design interactive systems to support and augment attention and memory rather than burdening them.
- For the purpose of user-interface design, the most important type of short-term memory is what psychologists call "working memory." It is not a separate memory store but rather everything we are *paying attention* to at a given time. The capacity of working memory—the number of concepts or ideas that we can attend to simultaneously—is five plus or minus one. Think of it as a small set of searchlights illuminating—a few at a time—ideas and concepts in a vast warehouse representing long-term memory.
- Working memory has severely limited capacity, so our attention is highly focused and selective. We miss most of what occurs around us because of this.
- Long-term memories record our experiences and thoughts. Thoughts and experiences trigger large patterns of neurons throughout the brain based on features. A memory is recorded as a collection of neurons that tend to fire in concert when stimulated by perceptions or thoughts. Psychologists divide long-term memory into semantic, episodic, and procedural.
- Unlike computer memory, reading information from human long-term memory can change it. It is also influenced by emotions. These factors make it error prone; it can easily lose or add details.
- VUIs spread information out over time, so they place a higher load on short-term memory than graphical user interfaces do.
- A user interface has modes if a control or user-action has different effects in different contexts. User interfaces that have modes burden working memory by requiring users to mentally keep track of what mode the system is in.
- Implications of memory and attention limitations for user-interface design:
  - Keep voice output brief so as not to overburden users' working memory. Augmenting voice output with screen output or indicator lights can reduce demands on working memory.
  - Avoid moded user-interface designs where possible.

- Minimize the number of calls to action on a page or screen. Avoid competing calls to action.
- Don't expect people to remember arbitrary facts or procedures over long periods. Provide memory aids and tools.
- Design user interfaces to attract users' attention where needed.