Chunking the Information Presented to Readers

A critique of Information Mapping

Geoffrey Marnell

There is a view—especially prevalent in the technical writing profession—that information should only be presented to readers in small quanta or chunks. On the face of it, this is a sensible view. A sentence composed of three or more clauses usually breaches the limit of cognitive lodgment. It is just too long for most readers to absorb. Only by atomising a conglomerate of ideas into discrete sentences—that is, by chunking our ideas—do we minimise the effort our readers need to exert to understand what we have written.

We also chunk our writing into paragraphs. A paragraph enables a writer to present an argument, discussion, analysis or whatever in logical and accessible units:

> "There are two purposes for paragraphs, the one logical and the other practical. The logical purpose is to signal stages in a narrative or argument. The practical one is to relieve the forbidding gloom of a solid page of text."[1]

Paragraphing is thus a form of chunking. Without it, readers would be presented with a form of *scriptio continuo*: undemarcated logic rather than undemarcated words. Such writing would hardly be inviting, nor likely to keep readers engaged. They would have to work out for themselves where each new topic began, causing them to drag anchor as they read. So chunking has been part of writing for a very long time. We chunk ideas into sentences, groups of related ideas into paragraphs, groups of related paragraphs into sections, and so on.

Traditionally, writers could choose the length of a sentence, a paragraph and a section. The amount of material they had about a particular topic, and the form of reasoning being adopted, might place limits on how long a paragraph or section might be. But these were limits imposed by logic and they varied from paragraph to paragraph. No one—or at least no one before the late 1960s—considered that there had to be an upper limit to the size of a paragraph or section no matter how much information the writer had to deliver. We might have deliberately avoided presenting readers with solid pages of text, but we did so without regard to any particular limit on the size of our chunks.

Today there are some who believe that we should limit our chunks to 7 ± 2 units of information, whether we are writing instructional materials, designing billboards or programming computer code. The source of this belief is usually attributed to an analytic literature review published by the American psychologist George Miller in 1956, the title of which is "The magical number seven, plus or minus two: Some limits on our capacity for processing information".[2] Many writers were introduced to the notion of a chunking limit through training in what was once a popular writing method: *Information Mapping*. The method is still taught and thus the so-called magical span of 7 ± 2 still holds a spell over many writers. But is this science or flim-flam?

In this paper we will consider Information Mapping in some detail. In particular, we will present evidence—both logical and empirical—that a chunking limit, whatever its size, is neither based on the research of George Miller, nor does it stand up to critical scrutiny. In other words, the so-called chunking limit at the heart of Information Mapping rests on weak foundations.

This paper is presented in three parts. The first part looks at the evidence adduced by the Information Mapping fraternity to support the 7 ± 2 chunking limit. The second part considers the view of the Information Mapping fraternity that reader comprehension requires that the 7 ± 2 chunking limit be applied to *every* part of a document. We show that, given the way people read, document-wide chunking is irrelevant. Readers simply do not notice the structure of a document when they are reading it And what they do not notice cannot affect their comprehension. Finally, we concentrate on paragraph chunking. Not only do we show that the Information Mapping limit of 7 ± 2 sentences on the size of a paragraph is unwarranted; we also provide empirical evidence that the number of sentences in a paragraph has little or no effect on a reader's ability to comprehend the paragraph. When all the evidence is assembled, there seems no reason for following the advice of Information Mapping and limiting each chunk of information presented to readers to 7 ± 2 sub-chunks.

1. N Hudson, *Modern Australian Usage*, OUP, Melbourne, 1993, p. 294

2. GA Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information", *The Psychological Review*, 1956, vol. 63, no. 2

Part 1: The 7 ± 2 limit—Science or bunk?

The Information Mapping method was developed by the American political scientist Robert Horn. According to Horn:

> "Writers should group information into small, manageable units … A 'manageable unit' of information is one consisting of no more than nine pieces of information … Rationale: Research suggests that people can best process and remember no more than seven plus or minus two pieces, or units, of information at one time … Therefore, a general guideline for a 'manageable unit of information' is one consisting of 7 ± 2 pieces (also referred to as the chunking limit) … Writers should create units of information that do not exceed the chunking limit. We should apply this limit at every level of a written document … By chunking information the writer improves the reader's comprehension and access and retrieval speed. Since readers can at best retain no more than 5 to 9 pieces of information in short-term memory, they comprehend material that has been 'chunked' more quickly and more completely."[1]

Furthermore:

> "… apply the Chunking Principle to:
>
> - sentences
> - blocks [that is, a group of 7 ± 2 sentences about a common topic, or a list or a table]
> - maps [that is, a group of 7 ± 2 blocks]
> - sections [that is, a group of 7 ± 2 maps] and
> - chapters [that is, a group of 7 ± 2 sections]."[2]

Also:

> "Remember that the Chunking Principle advises 7 ± 2 items in a list."[3]

And:

> "When your sentence is more than twenty words long, consider dividing it."[4]
>
> "[A] sentence [should] never be more than 30 words."[5]

The first question to ask is this: was Horn true to his own method in writing *Developing Procedures, Policies & Documentation*, the book from which these quotes are taken? It would seem not. The work includes numerous sentences of more than 30 words, has lists of more than nine items, and three chapters have more than nine sections or maps. But it would be churlish to put too much emphasis on such inconsistencies. The fact that the author does not follow his own method does not necessarily make the method flawed. To argue so would be to fall foul of an *argumentum ad hominem* fallacy.

What is a piece of information?

In Information Mapping, writing should be grouped into units of information—blocks, maps, sections, etc.—and a "manageable unit of information is one consisting of no more than nine pieces of information".[6] But what is a *piece* of information?

Imagine a map with seven blocks and with seven sentences in each block. This would fall within Horn's 7 ± 2 limit. Such a map would have 7^2 (or 49) sentences. So if a unit of information—in this case a map—cannot have more than nine pieces of information in it, then obviously a sentence is not what Horn means by a *piece* of information.

Now if a map can have no more than nine *blocks* and no more than nine *pieces of information*, it would seem that a block and a piece of information are considered the same thing. However, a block is also a "unit of information"[7], so if a unit of information is composed of pieces of information, a block cannot *exclusively* be a piece of information. Otherwise a block would always have just one element in it: one piece of information.

So it seems that for Horn what constitutes a *piece* of information varies from one unit of information to another. At the level of chapter, a piece of information is a section; at the level of a section, a piece of information is a map; at the level of a map a piece of information is a block and at the level of a block a piece of information is a sentence. But what is a piece of information at the level of a sentence?

This is an important consideration. The sentence is the fundamental, indeed necessary, building block in every document. Horn accepts that this is so: "The first and basic *unit of information* is the sentence."[8] But he also says that "Writers should group information into small, manageable units [and a] manageable unit of information is one consisting of no more than nine pieces of information".[9] So we would expect Horn to tell us what a piece of information *in a sentence* might look like. If the sentence is the first and basic unit of information, then the entire Information Mapping edifice balances or topples on the answer to this very question. Alas, Horn does not provide an answer. And for all its apparent importance, *sentence* does not even get a mention in the index of Horn's book. Neither does *piece of information* nor *information, piece of*. We have to work it out for ourselves.

1. RE Horn, *Developing Procedures, Policies & Documentation*, Info-Map, Waltham, 1992, p. 3-A-2
2. ibid. As we'll see, Horn did not apply his chunking limit to sentences, despite what he says in this quote. See page 3.
3. ibid., p. 10-5
4. ibid., p. 12-3
5. ibid., p. 12-2
6. ibid., pp. 3-B-2, 3-B-5 and 3-A-2
7. ibid., 3-B-1
8. ibid., p. 12-3. Emphasis added.
9. ibid., p. 3-A-2

A piece of information at the level of a sentence cannot be the same as a character or a word, for Horn allows up to 30 words per sentence.[1] A thirty-word sentence would then have at least thirty pieces of information in it, well above the specified limit of nine. Anyway, words like *the*, *a*, *an* and the like couldn't possibly be pieces of information. Perhaps no word on its own could be a piece of information.

After words, the next level of granularity in a sentence is a *phrase*. A phrase is a string of words that, although potentially meaningful when combined with other words, does not have a subject (a thing singled out for discussion) or a predicate (something said about whatever is singled out for discussion). Put another way, "a phrase is a group of words that act together as a unit within a sentence [but which] can't stand on [its] own and make a sensible message".[2] If a phrase has no subject nor can stand on its own to make a sensible message, then obviously it cannot be a piece of information.

The next level of granularity is the *clause* (that is, a string of words with a subject and a predicate). In fact, the clause is as far as we can go. After a clause, we have a sentence, and it would make no sense to say that the *fundamental* piece of information in a sentence is the sentence itself while at the same time allowing that a sentence, being a unit of information, can have up to *nine* pieces of information.

And now it should be clear why Horn did not extend his 7 ± 2 limit to the pieces of information that make up a sentence. The prospect of reading a nine-clause sentence would repulse most readers. Even a five-clause sentence would be indigestible to many readers, even those familiar with the topic. (A sentence composed of five independent clauses is equivalent to *five separate sentences* glued together with conjunctions and punctuation.) And this, no doubt, is why Horn opted instead for a word limit, not a pieces-of-information limit, on sentences: "an average of 20 words and never more than 30".[3]

The sentence is not the only unit of information that Horn excludes from his 7 ± 2 limit. If a chunk as large as a chapter can be a unit of information, there seems no reason why a book or even a large report should not also be considered a unit of information. (It is just a collection of chapters as a chapter is a collection of sections.) But Horn nowhere limits the number of chapters in a book or report.

Is there science behind Information Mapping?

Horn repeatedly states that his method is based on research. Here are just two of the many mentions:

> "Research suggests that people can best process and remember no more than seven plus or minus two pieces, or units, of information at one time."[4]
>
> "The … method implements research-based findings on how individuals process and understand information most efficiently."[5]

Despite being repeated a number of times in the book, no references to research to back up any particular claim are provided in footnotes or endnotes. There is a substantial "core bibliography" which, we are told, provides "complete citations for the research base on which the Information Mapping methodology was built"[6] and yet this bibliography includes many references of dubious relevance to Information Mapping. As one reads through it, a sense of quantity out-doing quality quickly intrudes. For example, there are references to generalist books such as *Statistical Methods* and *The Structure of Science*. How could such books be *specifically* relevant to Information Mapping? Including such general texts is about as silly as, say, Einstein, tendering a textbook on calculus in support of his general theory of relativity.

One also finds references in Horn's bibliography that seem hardly related to scholarly research of any kind, such as Whitlock's 1972 interview with Horn. Can a mere interview be classified as core research? There is also a reference to a 1945 article published in *Atlantic Monthly*, more a magazine than a research journal. Further, there are no immediately recognisable references to back up any specific Information Mapping principle, such as the chunking rule.

A fifteen-page bibliography might look impressive and suggest that much scientific rigor has gone into the recommendations in the book. But if the bibliography provides no help to the interested reader who wants to check how a recommendation is supported, it fails its very purpose. An author who makes a claim that is said to be based on research and yet provides no reference to that research is akin to the scientist who reports on some research but fails to explain to fellow scientists how the experiment was done and how it can be replicated. To expect the interested reader to read every item in a 15-page bibliography to verify that the accompanying text is soundly based on research is utterly unreasonable. Further, an "unpublished draft [of a] manuscript for [a] 1967 course at Harvard" along with three "unpublished proposal[s to the] US Air Force" are listed in *The Information Mapping Method: 30 Years of Research*, a 1999 publication of Information Mapping, Inc., the company Horn established to promote the Information Mapping method.[7] Can unpublished

1. ibid., p. 12-2
2. P Peters ed., *The Macquarie Student Writers Guide*, Jacaranda, Milton, 1989, p. 339.
3. RE Horn, op. cit., p. 12-2
4. ibid., p. 3-A-2
5. ibid., p. 2-3
6. ibid., p. B-1
7. Available from http://www.infomap.com. Viewed 21 January 2011.

course notes and commercial proposals really constitute research? Quantity over quality again?

The overlooked Miller

The American psychologist George Miller is widely credited with pointing out the potential significance of 7 ± 2 as a limit to various aspects of human cognition. In research conducted in the early 1990s, Miller's seminal paper—"The magical number seven, plus or minus two", referred to earlier—shared with two others the honour of being the most cited papers in "24 well-known introductory psychology texts [that] fully covered the field of psychology".[1] Horn tells us that he began developing Information Mapping in 1967, with the first version of his book on the subject published in 1976.[2] Given the excitement generated by Miller's paper, it seems odd that Horn does not include the paper in his own bibliography. Let's repeat that: Miller set the 7 ± 2 debate in train, his paper is much discussed and cited (and still is), twenty years later Horn expounds on the importance of presenting writing in chunks of 7 ± 2 and tells us that this is research-based—and yet neglects to mention Miller's paper. That is nothing less than odd. Just as odd is the fact that when Miller's and Horn's bibliographies are compared, not a single reference can be found in one that is also in the other.

The failure to acknowledge Miller, and the sources Miller quotes, becomes even more curious when we read, on the Information Mapping website, that:

> "The chunking limit is a guideline, *based on George A. Miller's 1956 research*".[3]

Can we assume that this is Horn's view too? Well according to Horn's own page on the Stanford University website, he has been chairman of Information Mapping, Inc. since 1987.[4] This is the company he established twenty years earlier to promote the Information Mapping method. The website of Information Mapping, Inc. is http://www.infomap.com, the very site from which the quote above was taken. Perhaps it is not stretching it too far to think that the chairman of a company would endorse the claims publicly made about the chief product of that company—especially when the product is the creation of the chairman. Thus it is not unreasonable to think that Horn did, and still does, think that George Miller's research supports the Information Mapping method. But does it?

Miller's magic number

For a start, let's make one thing clear. Miller did not do the research described in his justly famous 1956 paper about the limits of human cognition. He simply reported the results that other experimenters had published and tried to make sense of them (including the apparent fact that 7 ± 2 kept cropping up in various, unrelated studies as a limit to cognition, something he suggested, in the final paragraph of his paper, might be nothing more than "a pernicious Pythagorean coincidence"[5]). So to say that Information Mapping is "based on George A. Miller's 1956 *research*" is a little misleading. If it is based on Miller's 1956 *paper*—as opposed to any of Miller's own research—then it must be based on the research of those who Miller quoted. But, as noted in the last section, not one of the researchers Miller quoted is mentioned in Horn's bibliography. Let's put this down to sloppy citing on Horn's part and assume that he meant to say that he based Information Mapping on the research that prompted Miller to think that 7 ± 2 was somehow significant in cognitive psychology.

Miller's paper covers three distinct topics, all in some way related to memory. Since the Information Mapping website quoted above does not mention which topic or topics in Miller's paper form the basis of the Information Mapping method, let's look at each one.

The span of absolute judgment

Miller first considers a number of experiments in which subjects are asked to make absolute judgments about such stimuli as frequencies, loudness, saltiness, the size of rectangles, and the like. The objective was to determine how well humans can distinguish between differing levels of intensity of particular stimuli. All these experiments follow a similar pattern: subjects are exposed to a random sequence of varying stimuli after each stimuli had been given an identifying number. The subjects were then asked to repeat the numbers in the same order as the stimuli were randomly given. For example, a frequency of 100 Hz might be assigned the number 2, 6 000 Hz assigned the number 9, 8 000 Hz assigned the number 3, and so on. When exposed to tones at frequencies, say, of 6 000 Hz, 8 000 Hz and 100 Hz in that order, subjects were tested to see if they correctly responded 2, 9 and 3 respectively. The number of frequencies (or whatever stimuli is being used) is gradually increased until the number of correct responses drops to zero.

1. DW Gorenflo & JV McConnell, "The most frequently cited journal articles and authors in introductory psychology textbooks", *Teaching of Psychology*, vol. 18., no. 1, 1991, p. 8f. Miller's paper was cited in 22 of the 24 texts surveyed.
2. op. cit., p. i
3. See http://www.infomap.com/index.cfm/themethod/Mapping_FAQs (viewed 23 January 2011). Emphasis added. One might expect that a bibliography that includes a general textbook on statistics would include the very paper on which Information Mapping is based.
4. See http://www.stanford.edu/~rhorn/a/site/HornCV.html. Viewed on 13 January 2011.
5. op. cit., p. 96

In summarising an experiment conducted by Irwin Pollack in 1952 using tones of varying frequency as stimuli, Miller writes:

> "When only two or three tones were used the listeners never confused them. With four different tones confusions were quite rare, but with five or more tones confusions were frequent. With fourteen different tones listeners made many mistakes … The result means that we cannot pick more than six different pitches that the listener will never confuse. Or, stated slightly differently, no matter how many alternative tones we ask him to judge, the best we can expect him to do is assign them to about six different categories."[1]

This is a misrepresentation of Pollack's result. Pollack concluded that subjects could assign tones to only *five* categories:

> "… an informational transfer of approximately 2.3 bits is the maximum obtained. This is equivalent to perfect identification among only about 5 tones."[2]

Even so, the number of tones Miller quotes is 6, not 7—his magic number. Where, then, did Miller get his magic number? He got it by looking at a number of similar experiments—not just Pollack's—and taking the *average* span of immediate judgment (what he also calls the *channel capacity*, or the number of correctly identifiable *categories*, in Pollack's language):

> "… the channel capacities measured ranged from 1.6 bits for curvature to 3.9 bits for positions in an interval. Although there is no question that the differences among the variables are real and meaningful, the more impressive fact to me is their considerable similarity. If I take the best estimates I can get of the channel capacities for all the stimulus variables I have mentioned, the *mean* is 2.6 bits and the standard deviation is only 0.6 bit. In terms of distinguishable alternatives, this mean corresponds to about 6.5 categories, one standard deviation includes from 4 to 10 categories, and the total range is from 3 to 15 categories."[3]

Note that the channel capacities across a number of like experiments ranged from 1.6 bits (or 3 categories, when judging the curvature of lines) to 3.9 bits (or 15 categories when judging positions along a linear interval). Miller took what he thought was the mean value—2.6 bits of information—equated that to 6.5 categories, and then rounded it up to 7.

On the face of it, this looks like sloppy arithmetic. If the mean is 2.6 bits of information, it is much closer to 6 categories than 6.5: $2^{2.6} = 6.06$. To get a channel capacity of 6.5, the average number of bits would have to be 2.7 (as $\log_2 6.5 = 2.7$). To get Miller's magical number 7, the average number of bits would have to be 2.8 (that is, $\log_2 7$). In Miller's defence, however, he was looking for a *limit* or *ceiling* on absolute judgment and this limit obviously has to be a whole number. In that case, rounding up rather than down makes sense.

Let's ignore Miller's arithmetic and concentrate instead on the wide range of channel capacities observed for various stimuli. Miller was surprised that the range wasn't greater, but if channel capacity—that is, the span of absolute judgment—is relevant to writing, there is a world of difference between advising writers to limit, say, the number sentences in a block to 3 rather than 15. If Horn relied on Miller's consideration of the observed range of channel capacities, then he needs to explain why judging meaning (or whatever) is more akin to judging the frequencies of tones (where the channel capacity is 5, given Pollack's research) than judging positions along a linear interval (where the channel capacity is 15). To opt without reason for 7 is hardly scientific, given that Miller did not report any experiments on the absolute span of judgment when written material was the stimuli. Indeed, for all that Miller tells us, that span could fall outside the observed 3–15 range.

Anyway, the relevance of any span of absolute judgment to effective written communication is tenuous. Readers do not have to make the sort of comparative assessments of stimuli Miller considered in order to understand what they are reading. That is, understanding a piece of written material simply does not involve recalling the relative intensities of some semantic stimulus or other. There is, in other words, no parallel between naming tonal frequencies according to a provided legend and unravelling the meaning of a map, block or sentence. And, as we'll see a little later, Miller said that himself.

The span of attention

The next part of Miller's paper—entitled *Subitizing*—has to do with "the discrimination of number".[4] The experiments that Miller reports have long precedents. In the nineteenth century, the Irish mathematician William Hamilton and the English economist Stanley Jevons independently showed that if subjects were briefly shown marbles (or stones) in a box, they could easily remember the number they had seen up to about seven, after which

1. ibid., p. 83–4
2. I Pollack, "The information of elementary auditory displays", *Journal of the Acoustical Society of America*, 1952, vol. 24., no. 6., p. 748. Pollack's 2.3 bits of information is equivalent to $2^{2.3}$ correctly identifiable tones, which equates to 4.925. In general, the number of categories correctly judged (which Miller calls *channel capacity* and *span of absolute judgment*) is equal to 2^b, where b is the number of bits of information presented by the stimuli as a whole. Thus the number of bits associated with a channel capacity of, say, c is $\log_2 c$. For our purpose, we are only interested in channel capacity, not bits of information.
3. ibid., p. 86. Emphasis added.
4. ibid., p. 90

accurate recall fell away. Miller quotes more recent experiments that give the same result, in this case, dots flashed on a screen for 0.2 seconds at a time. An accurate count drops away dramatically after seven dots.

Could this be what Horn was basing Information Mapping on? It seems unlikely. No writer writes with the expectation that readers will only see what they have written for 0.2 seconds at a time (or two minutes, for that matter).

Span of immediate memory

The last part of Miller's paper is taken up with what he calls the *span of immediate memory* (also known as the capacity of our short-term memory). This may be what Horn based his method on given his repeated appeal to the limit of our short-term memory:

> "Since readers can at best retain no more than 5 to 9 pieces of information *in short-term memory*, they comprehend material that has been 'chunked' more quickly and more completely."[1]

In the two experiments that Miller cites, subjects were given several stimuli in succession and then asked to immediately recall them. The types of stimuli included binary digits, decimal digits, letters of the alphabet, letters plus decimal digits and monosyllabic words. The results:

> "With binary items the span [of immediate memory] is about nine [and it] drops to about five with monosyllabic English words."[2]

A span of 5 to 9 looks suspiciously like Horn's magic range of 7 ± 2. But just how relevant to comprehension is a measure of our ability to *immediately* recall a list of digits, words or whatever? Take one of Horn's maps, for example. A map can have 7 ± 2 blocks and each block can have 7 ± 2 sentences (or a list of 7± 2 items). Now consider the simplest map possible while keeping within the 7 ± 2 range: a map with five blocks each with a list of five items. Now try this experiment: on each of five cards write a list with five items in it. If you like, make it easier for yourself to remember the items by including only like items in each list (say the names of birds in one list, cities in another and so on). As you complete a list, turn the card over so that you cannot see what you have just written. When you have finished writing all five lists, try to recall, without looking at the cards, every item in all five lists in the order in which you wrote them. What are the chances of you correctly recalling all items? About the same as getting a money-back guarantee from a palm reader. What about recalling all the items in any order?[3] Only marginally better. You will probably recall the first few and last few items—what psychologists call the *primacy* and *recency* effects—but struggle with those in the middle. But if Miller's recall research is relevant to comprehension at the level Horn envisages, you should be able to correctly recall all 25 items, since they were in just five chunks (that is, five blocks). But if you can't correctly recall the items, you can't possibly have comprehended the material.

It might be retorted that a test of memory is not a test of comprehension, and that is true—in some cases. I might be able to recall the sentence "The work done was five ergs" but have no comprehension of what it means if I haven't studied physics. But in the case where you have listed the words to be recalled yourself, it is difficult to see what comprehension could be other than recall. Ability to recall at a later date, perhaps? Well that's hardly likely to be better than your ability to recall immediately after listing the items. Anyway, we are now straying far from Miller's paper, which we are told is the foundation for the Information Mapping method. Miller never examined delayed recall. Nor did he examine comprehension.

We poorly remember five lists of five items (and even shorter material) because, by definition, short-term memory is short:

> "Forgetting over intervals measured in seconds was found."[4]
>
> "… 30 seconds is ample time for forgetting to occur".[5]

This makes Horn's view that the limited capacity of our short-term memory should compel writers to apply the 7 ± 2 chunking rule at the *molecular* level—at the level of blocks, maps and sections—decidedly odd. By the time I've read even one seven-chunk block in, say, a map, 30 seconds is likely to have passed. If there are seven blocks in the map, many minutes may have passed by the time I have read them all. That is, as Baddeley says, "ample time for forgetting to occur". Further, a map could have over 50 items in it: 7 blocks each with a list of 7 items, plus all the headings. Given Miller's limit, my short-term memory will have been flushed out many times by the time I reached the end of that map. So what is the relevance of limiting a map to seven blocks? If 7 ± 2 applies only to *short-term* memory—and that is all that Miller said—why must it apply to chunks of material that could never be accommodated in a single frame of short-term memory?

1. RE Horn, op.cit, p.3-A-2. Also on page 12-2. Emphasis added.
2. Miller, op. cit., p. 92
3. It's not clear from Miller's paper whether the tests discussed were of serial recall (recall in the order given) or free recall (recall in any order).
4. LR Peterson & MJ Peterson, "Short-term retention of individual verbal items", *Journal of Experimental Psychology*, 1959, vol. 88, no. 3, p. 198. The experiments reported in this paper suggest that after about four seconds, correct recall drops to about 50%, and to zero after 18 seconds. See figure 3 on page 195 of the paper.
5. A Baddeley, *Working Memory*, Oxford University Press, Oxford, 1987, p. 10.

The capacity of our short-term memory might well be relevant to our ability to take in and comprehend material at the *atomic* level of a text—the clause or sentence—but at the *molecular* level—the level of blocks, maps, sections and chapters—its relevance is doubtful. If this is what Horn is claiming, then the onus is on him to advance some supporting research.

Flawed reading assumptions

Even if Miller had been right in thinking that the span of immediate memory is limited to 7 ± 2, the way this span was established—by testing *immediate* recall—and the span itself are quite irrelevant to the way people read. Suppose I want to know how to use time-shift on a personal video recorder and consult the accompanying user guide for instructions. Suppose, too, that I encounter a seven-step procedure (indeed, the number of steps is irrelevant for this argument). Am I meant to read each step and then recall them all before I can successfully complete the procedure? Of course not. Further, in a recall experiment such as those Miller considered, the subject cannot ask for the items to be repeated; in the real world a reader can go back and read an earlier step in a procedure, if they really needed to. The relevance of Miller's work is looking shakier and shakier.

Further, hierarchical chunking—from block to chapter—is unlikely to help most readers of many types of documents. Consider the types of documents Horn was primarily concerned with: procedures and policies. Most readers don't read whole chapters in user manuals and sets of work instructions, the types of documents where procedures and policies are prevalent. They dip into such documents when they want to learn (or be reminded of) how to do something in particular. They want to activate time-shift on their personal video recorder and the steps are not obvious. They might then scan the contents pages or index, or electronically search, for the topic of interest and then read *just that topic*. Perhaps in the effective life of a personal video recorder, the owner might consult the user guide a dozen times but never read it through in its entirety. Much to the chagrin of technical writers, a user guide is primarily consulted as a last resort: when the product it describes does not work as the user expected it to, or when the information needed cannot be got by asking someone else. And when it is, only a small part of it is consulted at any one time.

In Information Mapping, there is a special type of map for procedures and work instructions: the procedure map.[1] If most readers of instructional materials only dip into the materials to learn how to do a particular task, on each reading they are likely to be reading only one, perhaps two, maps. In which case they will not need any special guidance that might be given by the 7 ± 2 structure of a *section*. (Recall that a section is a group of 7 ± 2 maps.) Nor will they need any special guidance that might be given by the 7 ± 2 structure of a *chapter*. In other words, readers of procedures and work instructions are more interested in the trees than the wood. Indeed, most won't even see the wood.

Let's put this another way: will I *understand*, say, a procedure more quickly or more thoroughly if the chapter of which it is only a small part has seven rather than, say, ten sections? Unlikely. The *molecular* structure of the chapter will simply be unnoticed (and how can something I don't notice influence my degree of comprehension). Will I *find* that procedure more quickly with a seven-section structure than a ten-section structure? Hardly, given that I, like most readers, will go to the index or table of contents for help in finding a procedure. (And even if there were no index or table of contents, molecular chunking is unlikely to improve the speed with which topics are located. Indeed it may *impede* that speed, given that complex topics that should logically be kept together—and which readers would expect to find together—might have been split across chapters, and for no other reason than to avoid breaching Horn's chunking limit.)

All this puts paid to Horn's claim that:

> "By chunking information the writer improves the reader's comprehension and access and retrieval speed."[2]

The *molecular* structure of a document has no bearing at all on my ability to understand any *atomic* part of that document; nor does it necessarily improve how quickly I can find a particular part of that document.

This is more fully explored in the second part of this paper.

Is Miller relevant but memory not?

Horn's main point is that "since readers can at best retain no more than 5 to 9 pieces of information in short-term memory [writers are compelled to apply a chunking limit of 7 ± 2] at every level of a written document" (see page 7). Short-term memory, it seems, is important. It is *the* determining factor. We also noted that the website of the Information Mapping company of which Horn is the chairman states that "the chunking limit is a guideline, based on George A. Miller's 1956 research" (see page 4). As we've just seen, short-term memory is the very stuff of Miller's paper, especially its role in judgment, attention and recall. What, then, are we to make of the rest of the reference to Miller on the Information Mapping website:

1. RE Horn, op. cit., p. 4-B-1

2. op.cit., p. 3-A-2

> "The chunking limit is a guideline, based on George A. Miller's 1956 research, for creating information that people have to memorize. Documents do not have to be 'memorized', but maintaining these chunking limits aids in a reader's ability to process information."[1]

This is puzzling. If readers don't need to memorise material they read—which is true—then why must that material be limited to chunks that do not exceed the capacity of short-term memory? And why say that the chunking limit is based on Miller's research when Miller's research says nothing about what might aid a reader's ability to process information *other than information that has to be immediately recalled*. If we are not discussing information that needs to be immediately recalled—which we can't be if memorisation is not an issue—then we can't be basing our research on Miller. So where is the research that shows that even though memorisation is not required of readers, writers must "at every level of a written document" chunk their material in line with the constraints imposed by the capacity of short-term memory? Claiming Miller as an authority is looking a touch like fabrication.

But the very same web page on which we are told that "Documents do not have to be memorized" also tells us that:

> "Chunking… involves making the information digestible either for memorization or comprehension."[2]

So perhaps all that the first quote meant was that we don't have to memorise *whole* documents in order to understand their contents. That's too obvious to warrant discussion. But what about parts of a document? It's worth repeating here that Miller was concerned solely with *immediate* recall. Does anyone ever attempt to—or even need to—memorise a chapter, section, or even a map for *immediate* recall? Of course not. Does anyone even attempt to memorise a chapter, section, or map for *later* recall? Again, of course not. I might prefer to remember the main points in, say, a map so that I don't have to resort to the user guide in the future. But if my attempts at memorisation are not tied to the limits of my short-term memory— if I have time, that is, to rehearse and to mnemonically code what I want to remember—then Miller's magic number, pertaining as it does to *immediate recall*, is utterly irrelevant. And if the concern is to minimise the effort involved in rehearsing and mnemonically coding what I want to remember, then perhaps we should be applying the *minimum* chunk limit possible before recall errors occur: three (as noted in the next section). Miller was, you may recall, talking about the *limits of perfect recall*.

1. See http://www.infomap.com/index.cfm/themethod/Mapping_FAQs (viewed 23 January 2011).
2. ibid.

A parallel should reveal the absurdity of Horn's position, if such revelation is still needed. A listener does not need to memorise a melody, section or movement of a piece of music in order to appreciate it. (Otherwise we would never be enthralled by a piece of music we had never heard before.) Now let us suppose, for the sake of argument, that humans can only accurately hum 7 notes of any melody they hear for the first time when asked to immediately recall it. Does it then follow that composers should restrict to seven the number of notes in a melody, the number of melodies in a section, the number of sections in a movement, and so on? Again: of course not.

Research after Miller

In a 98-page article published in 2000, American psychologist Nelson Cowan reviewed the data then available on the limits of short-term memory. He noted that researchers post-Miller had found that short-term memory is limited to between three and five chunks. In summarising the masses of data he reviewed, Cowan concluded that there is a:

> "single, central capacity limit averaging about four chunks …"[3]

He also noted that:

> "[Miller's magic] number was meant more as a rough estimate and a rhetorical device than as a real capacity limit."[4]

Perhaps the Information Mapping fraternity might like to update their research instead of relying on Miller's survey of a handful of experiments conducted more than 50 years ago. If the span of short-term memory is only 4 ± 1 and if, as Horn claims, the span of short-term memory sets the chunking limit, then Information Mapping needs to be radically updated to bring it into line with current knowledge in cognitive psychology.

However, Cowan's reported limit is a limit on the recall of *unrelated* items (as was the limit reported by Miller). But when psychologists look at strings of *related* words—such as words in a sentence—the span of immediate memory is significantly greater:

> "Immediate memory for sentential material is typically substantially greater than span for unrelated words … Baddeley *et al.* … found spans of around five for unrelated words [in line with what Cowan reported] and 15 [words] for sentences."[5]

So if the span of short-term memory determines the chunking limit—as Horn contends—and if the span of short-term memory *for the sort of material that*

3. N Cowan, "The magical number 4 in short-term memory: A reconsideration of mental storage capacity", *Behavioral and Brain Sciences*, vol. 24, 2000, p. 87
4. ibid.
5. A Baddeley, *Working Memory, Thought, and Action*, Oxford University Press, Oxford, 2007, p. 143

writers primarily present to readers (namely sentences) is 15, perhaps the chunking limit should be raised to 15. Either way, Information Mapping has been left behind by research more recent than Miller's. A limit of 7 ± 2 is yesterday's guesstimate. Today it is 4 ± 1 for unrelated items and 15 words for sentences, the very entities that Horn claimed are the "first and basic unit of information" (see page 2).[1]

Then again, short-term memory, as we've seen, has little bearing on the optimum way to structure a document. It might be relevant to the construction of a sentence (or even a short paragraph). But it is irrelevant to the molecular structure of a document. The molecular structure is either unnoticed by readers or, if it is noticed, does not influence our ability to comprehend material at the atomic level.

What George Miller might say about Information Mapping?

Well-known author and freelance editor Mark Halpern wrote to George Miller in the mid-1990s when confronted with a workplace edict to limit the items in a list and the steps in a procedure to 7 ± 2. Halpern knew that Miller's name was associated with research on the limits of cognitive processing and that Miller had publicly complained about the unfounded conclusions some had drawn from his research. Miller replied to Halpern detailing one of those unfounded conclusions. In the 1970s, some local authorities had passed by-laws restricting the number of items that could be displayed on a billboard to 7 ± 2, using Miller's research to justify the laws. (It turned out that a group of landscape architects, funded by the big motel chains, had lobbied the authorities to introduce the law.) In his reply to Halpern, Miller said:

> "… the point was that 7 was a limit for the discrimination of unidimensional stimuli (pitches, loudness, brightness, etc.) and also a limit for immediate recall, *neither of which has anything to do with a person's capacity to comprehend printed text*."[2]

Perhaps the information mappers might wish to reconsider citing Miller's research as the basis for their chunking rule.

1. Note that Horn's omission of sentences from his 7 ± 2 chunking rule (see page 3) is not relevant here. We are only considering the raw capacity of short-term memory, which Horn uses to justify the chunking of blocks, maps, sections and chapters.

2. The entire thread between Halpern and Miller can be read at http://members.shaw.ca/philip.sharman/miller.txt. Emphasis added. Viewed 14 January 2011. In an email to me on 8 February 2011, Halpern confirmed the accuracy of his reported exchange with Miller. An email to Professor Miller at his last-known Princeton University address bounced.

Part 2: Document-wide chunking—Does size really matter?

In the first part of this paper, I dissected the claim by Robert Horn—the originator of Information Mapping, a writing methodology much loved by many technical writers—that research by American psychologist George Miller on the limitations of short-term memory shows that we should present information to readers in chunks of no more than 7 ± 2 sub-chunks:

> "Writers should … apply this [7 ± 2] limit at every level of a written document … By chunking information the writer improves the reader's comprehension … since readers can at best retain no more than 5 to 9 pieces of information in short-term memory …"[3]

The result of that dissection? Miller's research on the limitations of short-term memory in no way supports Horn's claim, and Miller himself was quoted as saying so.

The fact that Horn erred in basing his chunking principle on the limited capacity of our short-term memory does not on its own disprove that some chunking limit is necessary for comprehension. Perhaps, then, the chunking principle at the heart of Information Mapping is still appropriate even if the work of Miller cannot be adduced in support of it.

Let's assume, for he sake of argument, that there is some chunking limit. To what chunks in a document might it sensibly apply? Horn contends that it should apply to *every* level in a document:

> "Writers should … apply this [7 ± 2] limit at every level of a written document [namely] …
>
> - blocks [that is, a group of 7 ± 2 sentences about a common topic, or a list or a table]
> - maps [that is, a group of 7 ± 2 blocks]
> - sections [that is, a group of 7 ± 2 maps] and

3. "The chunking limit is a guideline, based on George A. Miller's 1956 research". See http://www.infomap.com/index.cfm/themethod/Mapping_FAQs (viewed 23 January 2011).

- chapters [that is, a group of 7 ± 2 sections]."[1]

Just how plausible is this claim?

First let's note that there is nothing in Horn's claim that "chunking information … improves the reader's comprehension" to suggest that he is using the word *comprehension* other than in its common-or-garden dictionary meaning:

> "**comprehension** *noun* 1. the act or fact of comprehending"
>
> "**comprehend** *verb (t)* 1. to understand the meaning or nature of"[2]

Thus Horn's claim is that chunking improves a reader's ability to *understand* what they are reading. Stripped of whatever academic mystique swirls about the concept of *comprehension* should help us see that the relationship between chunking and *comprehension-as-understanding* is a tenuous one indeed. For if such a relationship exists, then it plainly cannot apply to *all* types of writing. A novel, for instance, is largely unchunked. It will have paragraphs (that is, blocks) and it may have chapters. But it won't have maps and very few novels have sections. And of the chunks a novel does have, no limit is adhered to by the author. A chapter could have hundreds, possibly thousands, of chunks. Does this lack of rigorous chunking in the form required by Information Mapping mean that we do not *understand* what we read in novels? Of course not.

Component	Max. number
document	–
… chapter …	unstated
… section …	7 ± 2
… map …	7 ± 2
… block …	7 ± 2
… sentence …	7 ± 2
… word …	30

Figure 1: Information Mapping: chunking hierarchy

1. RE Horn, *Developing Procedures, Policies & Documentation*, Info-Map, Waltham, 1992, p. 3-A-2
2. Macquarie Dictionary, www.macquariedictionary.com.au. Viewed 11 May 2011.

The way a reader might read

Let's leave novels and their kin to one side and consider instead user guides and policy manuals, the sort of documents Horn is primarily concerned with. Now a reader might tackle such a document in the following ways:

1. read some part of it (a section, map or block) without first paying any attention to the number of parent or grandparent chunks
2. read some part of it (a section, map or block) after paying attention to the number of parent or grandparent chunks
3. read some part of it without first paying any attention to the number of sibling chunks
4. read some part of it after paying attention to the number of sibling chunks
5. read some part of it (a chapter, section or map) without first paying any attention to the number of child or grandchild chunks
6. read some part of it (a chapter, section or map) after paying attention to the number of child or grandchild chunks

These six scenarios cover all logical possibilities.

Looking upwards

Take scenario 1: reading without first noticing any higher-order structure. This covers the vast majority of cases. Most readers of technical documentation will go to the index or table of contents for help in finding the chunk—a block, map or section—they want to read, go to that chunk and start reading it. The number of chunks in the *chapter* of which the block, map or section is a part will pass unnoticed. For example, if I want to learn how to set preferences in Adobe Photoshop, the index points me to a particular page. I go to that page and start reading. I don't even notice that this section (which is a *map* in Information Mapping terminology) happens to be in chapter 1; and I don't bother—why would I?—to flick through the chapter looking at the other chunks in it. I'm interested only in learning how to set preferences. Now if I don't pay any attention to the size of the parent or grandparent chunks—the sections and the chapters in this particular example—how can the size of those chunks affect my comprehension of what I am reading?

This way of reading user guides—indeed, of reading most informational texts—is widespread. Few readers, if any, read whole chapters in user guides. They dip into such documents when they want to learn (or be reminded of) how to do something in particular. They want, say, to activate time-shift on their personal video recorder and the steps are not obvious. They might then scan the contents pages or index, or electronically search, for the topic of interest and then read *just that topic*.

Perhaps in the effective life of a personal video recorder, the owner might consult the user guide a dozen times but never read it through in its entirety. Much to the chagrin of technical writers—who might spend many months writing a user guide—user guides are primarily consulted as a last resort: when the product it describes does not work as the user expected it to, or when the information needed cannot be got by asking someone else. And when it is, only a small part of it is consulted at any one time.

Let's explore this scenario a little further. Suppose that A is given version 1 of a user guide and B—of equal intelligence and experience—is given version 2. Suppose further that the only difference between the versions is that version 2 has more sections in chapter 1. A and B are then asked to read, say, a map in chapter 1, a map that is identical in both versions, and to do so without first counting the number of sections in the chapter. Must B find the map more difficult to comprehend than A? Surely not. Something that we do not look at in a document cannot possibly affect our comprehension of something that we do look at. And this will be the case regardless of the number of sections in the unlooked-at parent or grandparent chapter. It is also the case regardless of the type of chunk, even as far down the document hierarchy as the sentence. My understanding of "1. Sort the changes in postal code order" (to borrow an example from Horn) is in no way compromised by the number of unobserved sections or even maps in the grandparent chapter. In other words, a chunk-size limit is entirely irrelevant in this majority-case scenario.

Consider scenario 2: reading after noticing the size of the parent or grandparent chunk. First, note that this is an unrealistic scenario. How many of us, when we want to follow a procedure, first look at how many *sections* there are in the user guide? Few if any. How many of us, when we want to read a section of a policy document, first count the number of chapters in the document? Again, few if any. Still, for the sake of argument, let's explore the logic here. Suppose I read, in version 1 of a manual, "1. Sort the changes in postal code order". This is a one-sentence block in a procedure map. Suppose further that there are n sections in the grandparent chapter. Twelve months later, a new version of the manual is released and now that chapter has more than n sections in it. This is a fact I happen to notice. Indeed, just for fun, I count the number sections in the chapter. Now I go back to the procedure I read twelve months earlier and once again encounter "1. Sort the changes in postal code order". Is that step now more difficult to comprehend—and more difficult simply because of the extra sections in the chapter)? Hardly. And again, the size of n is irrelevant. Or returning to our earlier example, will B find the map to read more difficult to comprehend than A simply because B *notices* that version 2 of the user guide has more sections in it than version 1. Again, unlikely. In other words, a chunk-size limit is also irrelevant in this scenario.

Let's approach this scenario from a different angle: suppose that I want to understand a map in a section and I notice that there are n other maps in the section. Suppose further that the map I want to understand is very short: it has, say, just two blocks in it and each bock is composed of just two single-clause sentences. If a chunk-size limit did affect comprehension of a child block, we would have to say that, although this map should be very easy to understand—it has only four sentences in it—it would become difficult to understand for values of n above that limit. In other words, readers would then find those four sentences difficult to comprehend, and solely because the section has 15, 20, 50 or whatever *other* maps in it. This is simply nonsensical. Simplicity is simplicity whatever may clutter around it.

Looking sideways

Consider now scenario 3: read some part of a chunk without first paying any attention to the number of sibling chunks. Suppose I start reading a procedure without noticing that it has n steps in it. Perhaps there are 5 steps on one page and $n - 5$ on the next, with the latter not observed until I turn the page. Can it possibly be that my comprehension of any one of the first 5 steps is somehow influenced by the number of steps on the next page? Will I better understand "1. Sort the changes in postal code order" if there happens—unbeknown to me—to be just 2 steps on the next page rather than 20? Again, an affirmative answer seems highly unlikely. And again, the size of the n seems entirely irrelevant. Further, the same applies regardless of the type of chunk we consider. For example, my ability to comprehend a section is not compromised by the number of unnoticed sibling sections. Thus a chunk-size limit is irrelevant in this scenario too.

Consider now scenario 4: read some part of a chunk after paying attention to the number of sibling chunks. Again, this is a unrealistic scenario. How many of us count the number of steps in a procedure before working our way through it? Few if any. How many of us count the number of maps in a section before tackling a map? Again, few if any. Still, let's explore the logic of this scenario, if only to cover all bases.

Suppose that I read, in version 1 of a manual, "1. Sort the changes in postal code order" as the first step in an n-step procedure. Twelve months later, a new version of the manual is released and the very same procedure now has $n + p$ steps. Suppose that I notice, before I read the updated procedure, that the number of steps is $n + p$. Will I now struggle to understand "1. Sort the changes in postal code order" for certain values of n and p? Again, an affirmative answer

seems absurd. Once again, a chunk-size limit is irrelevant.

A parallel argument to one adopted earlier is also relevant here. Suppose that I want to understand a section in a chapter and I notice that there are *n* other sections in the chapter. Suppose further that the section I want to understand is very short: it has just one map with, say, two blocks and each bock has just two single-clause sentences. If a chunk-size limit did affect comprehension, we would have to say that, although this section should be very easy to understand—it has only four sentences in it—it would become difficult to understand for values of *n* above that limit (and where the value of *n* is noticed by the reader). In other words, readers would find those four sentences difficult to comprehend, and solely because the chapter has 15, 20, 50 or whatever *other* sections in it. This is simply nonsensical.

If a chunk-size limit did affect comprehension, how does it do so? Do we fully comprehend all the sub-chunks up to the limit and then struggle with those beyond it? Or does the breach of the limit somehow spread its tentacles throughout the entire chunk and affect our comprehension of every sub-chunk in it? Let's assume, for the sake of argument, that the chunk-size limit is 9. Will I fully comprehend the first 9 steps in a procedure and then struggle with the remaining 2? Or will I struggle with all 11 steps? Neither case seems plausible. A step such as "30. Select **Exit** from the **File** menu" would be comprehendible to most, probably all, of the intended audience regardless of how long the procedure is.

Looking downwards

Consider now scenarios 5 and 6. If we read a chunk in its entirety—as we would have to in order to claim in good faith that we understood it—then we cannot but read the child and grandchild chunks. It doesn't matter whether we notice the granularity of the sub-structure, that is, the number of sub-chunks: to comprehend a chunk we need to read the sub-chunks. But more than that, we need to understand the sub-chunks. For instance, I can't be said to understand a chapter without understanding the sections that make it up. And I can't be said to understand a section without understanding the maps that make it up. And so on as we move down the document hierarchy. So the question of comprehending a chunk—any chunk—comes down to comprehending the atomic components in the hierarchy, that is, the components that are not themselves chunks. In Information Mapping, such an atomic component—what Horn calls the "first and basic unit of information"—is the sentence. From an Information Mapping perspective a sentence is like a quark: it has no chunkable components. And Horn *explicitly* refuses to apply the 7 ± 2 chunking limit to sentences.

Now if understanding a chunk means nothing more than understanding the sentences that comprise it, then scenarios 5 and 6 are reducible to scenarios 1–4 *applied at the sentence level*. But recall that we used a sentence—*Sort the changes in postal code order*—in each one of scenarios 1–4. We showed that this sentence could be understood independently of the number grandparent or parent chunks, and the number of sibling sentences. This was so even if we noticed the number of parent or sibling chunks. Now there is nothing special about this sentence that gives it a better chance than any other of being independently understood. Hence we can generalise: if a chunk is understood only if we understand the sentences that comprise it, and if a sentence—any sentence—can be understood irrespective of the size of any chunk in the document of which it is a part, it follows that comprehension is in no way impaired by the size of a chunk regardless of where that chunk is in the document hierarchy.

To sum up: contrary to the claims of Information Mapping, document-wide chunking is not a necessary pre-condition of comprehension.

What does it all mean?

It follows from our six scenarios—scenarios that cover all possible reading practices—that comprehension does not impose any limit on the size of the chunks in a document. Comprehension may well impose a limit on the size of sentences—based on the capacity of our short-term memory—but beyond that, size does not matter. So it would be false to claim that there are too many sub-chunks in a chunk for that chunk, or any other chunk, to be comprehended.

Residual memory: an epiphenomenon of quantum incredulousness?

It might be retorted that while it is true that to understand a chunk you must understand the sentences in it, it is still possible that you could understand all the sentences but fail to understand the chunk. In other words, understanding the sentences in a chunk is a necessary *but not sufficient* condition for understanding a chunk.

With regards to procedure writing, this claim is highly improbable. Suppose I work my way through every step in a 20-step procedure and, in doing so, I achieve what I set out to achieve: reconcile a bank account, set preferences in Adobe Photoshop, replace a drive belt on a robot, or whatever. Can it make sense to say that although I understood each step—how could I not if I completed them?—I still didn't fully understand the procedure? I might not have understood *why* I had to do a particular step, but that is not the same as not understanding the step (for to understand a step is to know how to do what it commands). In other words, to understand a procedure is to understand the steps in it, and when

you understand all the steps in it you understand the procedure. There is no meaning that the map (in this case a procedure) has over and above the collective meanings of the constituent blocks (in this case, sentences). There is no semantic by-product, no epiphenomenon, no ghost in the machine.

So a procedure is comprehended once we've comprehended the steps in it. But what about conceptual material (such as the material you are reading now)? Might it be possible, say, for someone to understand every sentence in a multi-sentence paragraph but still fail to understand the paragraph?

In one sense this is possible, but that sense is irrelevant to our argument. If we are to keep the focus on the potential relationship between comprehension and *paragraph length*, we need to discard those reasons why a paragraph might be baffling that are obviously unrelated to paragraph length. For example, a paragraph that is poorly formed—one that is lacking the features readers customarily expect to find in paragraphs—can baffle a reader even if every sentence in it is understood perfectly. For example, I might find a paragraph baffling if:

- there is no topic sentence or no clear indication of what the paragraph is about
- some of the sentences in it seem unrelated (so that the paragraph lacks cohesion)
- the way sentences in it are linked is odd and distracting
- it contains a logical fallacy
- it contains sentences that contradict one another or contradict what was written elsewhere
- it is unclear whether the paragraph is about a new topic or continuing the discussion from the previous paragraph

But these are all faults with the way the paragraph has been constructed (faults in the sense that they do not give readers what they conventionally expect from a paragraph). Such faults lead to what might be called a *poorly formed* paragraph. On the other hand, a *well-formed* paragraph is one that meets readers' expectations: it has a distinct topic sentence (or its purpose is clear), all sentences are related and appropriately linked, the logic is impeccable, and so on.

Now we can recast the question at hand as this: is it possible that, *in a well-formed paragraph*, you can understand the meaning of the sentences in it but still fail to understand the paragraph?

To answer that question in the affirmative is to assume the existence of some meaning over and above the aggregate meaning of the sentences. Let's call this epiphenomenon *residual meaning*. We can then ask how are we to identify or detect residual memory.

For a start, it is not intuitively clear what residual meaning might be and how it might present itself. Perhaps it is akin to drawing a conclusion that was not stated from a set of premises that were stated. We read and comprehend the premises (the sentences) and then, by the exercise of our logical faculties, deduce the conclusion (the residual meaning). But that doesn't seem to match what occurs when we read. We do not appear to be inferring new meaning or new facts—meaning or facts unstated by the author—from every paragraph we read.

Could residual meaning be tied up with the realisation a reader might have of how all the sentences in a paragraph fit together, contributing to the discussion of the subject that the topic sentence introduced? But if the writer has written a *well-formed* paragraph, the linking of sub-topics and the overall paragraph cohesion should be apparent to the reader. This is typically achieved through linking words. Common linking words or phrases are *therefore*, *so*, *hence*, *accordingly*, *it follows then*, *as a result*, *in contrast*, *at the same time*, *however*, *further* and *for the same reason*:

> Doxycycline at 250 mg killed 90% of the bacteria. *In contrast*, penicillin at the same dosage killed at best 60%. *However*, penicillin at 500 mg killed all the bacteria. *Therefore*, our recommendation …

A pronoun can also link one sentence with another.

> Melanoma is a particularly aggressive form of skin cancer. *It* kills 2000 Australians every year.

Such linking is explicitly added by the writer. The paragraph cohesion subsequently attained—and noticed by the careful reader—is formed from the meanings of the sentences in the paragraph, including the linking words, pronouns and the like. It is not an epiphenomenon that arises as a by-product of something else that the writer has done (such as a conclusion that could have been drawn from premises but wasn't). The writer does it, the reader sees it—and sees it in the words that contribute to the meaning of the sentences. There is no spooky quantum-like phenomenon that mysteriously arises from the meanings of words that is not itself just the meanings of the words.

Could residual meaning be the mind's summary of what has just been read (or the knowledge that has been extracted from the words just read)? For a start, most of us don't summarise as we read. We might summarise if we were studying for an exam, but study is a minor part of our total reading life. When we read, say, an annual report, we do not read it with the expectation that we are going to be examined on parts of its contents. We don't even read it with the expectation that we need to remember what we have read. If perchance we needed to recall what was in a report, we can simply re-read the relevant parts. There would be too much unnecessary effort—and subsequent cognitive clutter—if every time we read,

we read in the hope of permanently remembering every fact.

Further, what might there be in a summary that is not also in the words being summarised? The knowledge we gain from reading is indeed separate from the sentences we read: the latter are aggregates of glyphs on paper or on a screen; the former a neurological loop in our long-term memory. But this existential independence doesn't mean that a neurological loop is *semantically richer* than the sentences that sparked its existence. Indeed, it is likely to be semantically poorer. Otherwise we would all have perfect memories.

It is difficult to see, then, what residual meaning could possibly be. In informational writing (as opposed, say, to poetry), we do not write with the expectation that the reader will get more from what we've written than what the sentences we've written denote. Indeed, that could almost be a definition of *poor* informational writing. If a writer's words cannot tie down the meaning of an otherwise well-formed paragraph, then either the writer is inept or the act of written communication is everywhere and forever mired in ambiguity and vagueness. On the other hand, if a writer's words do tie down the meaning of a well-formed paragraph, then there is no need for the concept of residual memory. The words say it all.

Further, it is difficult to see how residual meaning might present itself, especially if it is tied to a chunking limit. Does such meaning appear at some stage up to the limit—but when?—and then attenuate or disappear once that limit is breached? Our reading experiences do no seem to match this model of a waxing and waning meaning.

The concept of residual meaning is looking somewhat elusive, more at home, perhaps, in the metaphysical than the physical realm. There are certainly no hints in Horn's work that might help us identify and detect this mysterious form of meaning. Perhaps, then, Ockhams Razor can legitimately be wielded: of two competing hypotheses, the one that requires the fewest assumptions is the one more deserving of belief. And that hypothesis is the one that does not rely on the existence of residual meaning in a well-formed informational paragraph. In other words, the better hypothesis is that, in the context of semantic chunking, the whole is no more than the sum of the parts.

Returning to our main point:

Premise 1: to understand a chunk is to understand the sentences that make up that chunk, and to understand the sentences that make up a chunk is to understand that chunk. And this applies regardless of the chunk we are talking about, since our comprehension of any chunk can be reduced to our comprehension of the sentences in each of the lowest-level chunks.

Premise 2: readers can understand sentences regardless of how many sibling, parent and grandparent chunks there happens to be in the document they are reading.

Conclusion: Comprehension is entirely independent of chunk size. Thus the stress Information Mapping places on chunk size appears misplaced—and would be so *regardless of the chunking limit proposed*.

Note that in claiming that to understand a paragraph is nothing more than to understand the sentences that compose it we are not saying that each sentence *on its own* is necessarily understandable. We often use pronouns in writing to avoid endlessly repeating a noun. A pronoun can refer to a noun in the same sentence or to a noun in a previous sentence. When it refers to a noun in a previous sentence, then what it is referring to will not be apparent to a reader unless they had also read the previous sentence. Thus the sentence *It kills 2000 Australians every year* will, if read in isolation, be difficult or impossible to understand. It needs to be read in conjunction with the previous sentence: *Melanoma is a particularly aggressive form of skin cancer*. However, our point is that the meanings of the sentences in a well-formed paragraph, *when taken together*, constitute the meaning of the paragraph.

Addition or diminution?

So far we have considered what might be *added* to our understanding by grouping a set of individually understood sentences into a paragraph. The answer appears to be nothing. If so, our understanding of a paragraph is nothing more than our understanding of the sentences from which it is composed, a conclusion that throws doubt on Horn's belief that comprehension requires chunking.

However, comprehension is sometimes measured by a reader's ability to *recall* what they have read (which, incidentally, takes us beyond the dictionary definition of *comprehension* as *understanding*). Perhaps what Horn had in mind is that chunking improves a reader's ability to correctly recall what they have read: the bigger the chunk they have to read, the more difficult it is for a reader to recall that chunk. This appears to match the folk wisdom that the longer the paragraph, the more difficult it is to understand.

So the fact that our understanding of a sentence is not impeded by the number of sibling or parent chunks might not be the full story. Perhaps our ability to recall the contents of a paragraph might diminish as its size exceeds a certain limit (in which case, size would mater). Perhaps, then, we should be looking not at what is added as sentences accumulate in a paragraph, but what is lost.

But what does *recall* mean? In the experiments Horn bases Information Mapping on—the

experiments reported by George Miller in 1956—recall meant *memorisation*: that is, *perfect* recall. But no reader is expected to memorise what they read. We noted this in the previous part of this paper and noted, too, that this point is acknowledged on the main Information Mapping website. Thus we could accept that recall becomes less accurate the longer the paragraph without accepting that this imposes any obligation on the writer to limit the lengths of paragraphs. If readers don't read to memorise, then why should we tailor our writing in ways that might help them memorise?

However, some readers might want, at least some of the time, to remember what they read. They might not need to memorise material word-for-word, but they might want to be able to recall the salient facts that have been presented. Someone studying for an exam would fall into this group. Perhaps, then, writers should write in ways that meet the needs of such readers. And since we rarely if ever know whether a potential reader will want to *study* what we write, perhaps it is best simply to adopt an over-arching principle of always writing in ways that meet the needs of the studious reader. Since the studious reader wants to maximise recall—at least recall of salient facts—then perhaps we should be looking for a limit beyond which maximum recall of salient points begins to decline. In other words, although we might understand every sentence in, say, a fifteen-sentence paragraph, our ability to recall the salient points in it is compromised by its length being beyond what we might call the *salient recall chunking limit*. And, on the face of it, the *salient recall chunking limit* should be greater than any corresponding Miller-based limit (a limit that could be called the *perfect recall chunking limit*).

In the next part of this paper, I'll present research that shows that a salient recall chunking limit is impractically low. More importantly, the same research shows that it is *conceptual density* and *familiarity* more so than paragraph length that determines salient recall. I'll also present research that shows that our understanding of a paragraph appears unrelated to the number of sentences in it. Our conclusion will be that paragraph chunking is important, but that the number of sentences in a paragraph should be the writer's least concern when deciding where to chunk. Other issues are far more important.

Does paragraph length affect comprehension?

This is the last of three papers looking at claims made by the Information Mapping fraternity. In the first part of this paper, I dissected the claim by Robert Horn, the originator of Information Mapping, that research by American psychologist George Miller on the limitations of short-term memory shows that we should present information to readers in chunks of no more than 7 ± 2 sub-chunks. I showed that Miller's research did not state or imply this, and quoted Miller himself saying that none of his research has any bearing on documentation and comprehension. Thus the appropriated intellectual foundation of Information Mapping's 7 ± 2 chunking limit does not stand up to scrutiny.

In the second part of this paper, I looked at the way people read texts and concluded that the size of textual chunks—grandparent, parent and sibling chunks—does not affect our ability to understand any particular chunk of text. Chunks in the vicinity of our reading are simply unnoticed—or not analysed for size—by readers. This put paid to Information Mapping's claim that we should apply the chunking limit—the limit erroneously based on George Miller's research—to every part of a document (from chapter down to paragraph).

In this paper, I continue to investigate the claim that comprehension requires some chunking limit. Horn's limit (7 ± 2) might be spurious, but some limit, surely, must apply if comprehension is not to be compromised. Or must it?

Surely something is lost as size increases?

In the last part of this paper, I noted that comprehension is often measured by a reader's ability to *recall* the salient facts in what they have read. Perhaps what Information Mapping has in mind in claiming that "in chunking information the writer improves the reader's comprehension …"[1] is that chunking improves a reader's ability to correctly recall what they have read. The bigger the chunk they have to read, the more difficult it is for a reader to recall the salient facts in the chunk. This matches the folk wisdom that the longer the paragraph, the more difficult it is to understand:

> "For general purposes, paragraphs from 3 to 8 sentences long are a suitable size for developing discussion, and some publishers recommend an upper limit of 5/6 sentences."[2].

1. RE Horn, *Developing Procedures, Policies & Documentation*, Info-Map, Waltham, 1992, p. 3-A-2.
2. P Peters, *The Cambridge guide to Australian English usage*, Cambridge University Press, Cambridge (UK), 2007, p. 595.

> "Paragraphs should be kept short wherever possible."[1]

Note, for a start, that if we were asked to read a paragraph *for recall*, we would read it differently than if asked just to read it. (By *recall* here I mean the ability to repeat the main points in what one has read, not the ability to accurately regurgitate what one has read in the order that it was read.) To read a text for recall is to study that text, and to study a text is not the same as reading a text. Our reading strategies are quite different. Studying involves mental rehearsal, repetition and summarisation, and possibly note-taking. Through such strategies we improve our chances of correct recall. But these strategies remain idle when we read a newspaper, novel, article or report. Here we are happy to absorb each sentence as it is presented to us free of the mental gymnastics required in, say, preparing for an examination. A student might take hours to study a section in a textbook that the casual reader might read in 30 minutes.

In other words, we must be careful not to confuse *understanding* with *knowing*. I can understand how to replace the drive belt on a robot by understanding the steps in the procedure and replacing the belt. But if I wasn't expecting to be *tested* on my knowledge of the procedure, I might subsequently fail a closed-book exam that asked me to outline the steps required to replace the belt. To repeat: our reading styles when we are *studying* differ markedly from our reading styles when we are just reading for immediate throw-away facts. We might call the two styles of reading *studious reading* and *transient reading*. Students mostly engage in studious reading—at least when studying—while the rest of us engage in transient reading (at least when reading informational texts). We are reading simply to gain immediate information: for one-off understanding or one-off use. And a reader can understand a sentence without necessarily being able to recall it, or its informational content, hours, days or weeks later. (Did you understand the sentences in this paragraph? Make a note to yourself to try to recall its salient facts tomorrow.)

For the sake of argument, let's suppose that Information Mapping's chunking limit was found to be necessary to maximise *recall* (a point that has not been proved, incidentally). We have already established that the chunking limit is not necessary for *understanding*, but let's assume that it is necessary for recall. If so, what reason might there be to apply the chunking limit to "any business writing task and any type of document"[2] and "at every level of a written document".[3] Why force writers *at all times* to adopt a methodology supposedly best suited for study and recall when most readers don't read documents for study and recall? Most of us don't want to clutter our minds with one-off facts and throw-away information, especially information we know we can readily access again if we need to (by, for example, rereading a procedure or report). Does any one *study* a procedure? No. We read through it step by step (or perhaps just read those steps that we are unsure of). Likewise, we don't *study* a policy document, annual report or business memorandum. We simply read in the hope of understanding what it is we are reading but without the burden of needing to recall what we read. In a word: if most readers engage in transient reading, why force authors to write in a way that will assist studious reading (assuming that that is the case)?

It could be retorted that even though most of us only engage in transient reading, we should always assume that there will be *some* readers who engage in studious reading and thus we should—if we are to write with respect for all our readers—write with them in mind. The transient reader would not be burdened by disciplined chunking, but the studious reader would be helped. Fine if it were true that some chunking limit favours the studious reader without disadvantaging the transient reader. But is there research to show that?

Let's start with logic

Information Mapping assumes that the sentence is the fundamental unit of information:

> "The first and basic unit of information is the sentence."[4]

Linguists would dispute this. Indeed, a moments reflection will reveal that a unit of information more fundamental than the sentence is the *clause*. A clause is a string of words with a *subject*—something that is singled out for discussion—and a *predicate*—something that is said about the subject. A sentence can be a single clause—such as *The experiment was a failure*—or an amalgamation of several clauses, as in *The cyclone has passed and not a single building is still standing*. In the first sentence, there is a single unit of information: the experiment failed; in the second there are two units of information: there was a cyclone and it destroyed all the buildings. Similarly, a three-clause sentence offers three units of information, as in *Since we missed the last bus and there were no taxis about, we walked home*.

Suppose that there is some sentence-based chunking limit to paragraph comprehension and suppose further that it is 9 (in line with Information Mapping's 7 ± 2, although the actual limit is not

1. ISO/IEC 26514, Systems and software engineering—Requirements for designers and developers of user documentation, ISO, 2008, p. 94.
2. Horn, op cit., p. 2-1.
3. ibid., p. 3-A-2
4. ibid., p. 12-3.

pertinent to this discussion). Suppose further that you have two paragraphs: *A* and *B*. *A* is composed of 7 sentences each of 3 clauses (making 21 units of information in total), and *B* is composed of 10 sentences each of 1 clause (making 10 units of information in total). On the face of it, we seem to have a paradox: the paragraph that meets the chunking limit (*A*) has twice as many units of information in it—and thus you would expect it to be harder to comprehend—than the paragraph that exceeds the chunking limit (*B*). How can this be so?

Of course, it won't be the case if you assume transient reading. But we are discussing *studious* reading here, where readers can adopt strategies of rehearsal, repetition and summarisation to remember the salient points, unencumbered by a time limit. Thus it is possible for someone to recall all 21 units of information in paragraph A.

But in accepting this we must also conclude that the chunking limit, whatever it might be, is irrelevant in studious treading. If I am motivated to learn, there is no time limit on my study and I adopt effective study practices, there seems no limit to the number of units of information I could recall. How else does one pass an exam given the convoluted, verbose, multi-clause stew commonly served up in academic textbooks? How else do people memorise π to a thousand and more decimal places?

So, either a chunking limit leads us to a paradox—more is easier than less to comprehend (transient reading)—or it is irrelevant (studious reading).

What does the research show?

For the sake of argument, let's persevere with the idea that there must be a sentence-based chunking limit for paragraph comprehension *understood as the recall of salient points*. For a chunking limit to be useful in ensuring maximum recall, there must be a sentence-count below which recall is near enough to perfect and beyond which recall begins to deteriorate despite how studious the reader has been. That is what is meant by a *limit*. As Miller reported in experiments on the span of immediate memory, the immediate recall of, say, a list of arbitrary digits is usually close to perfect up to about 7 digits and then starts to fall away. Is there a similar limit (*n*) in relation to paragraphs such that information is correctly recalled in paragraphs of up to *n* sentences but not so in paragraphs of more than *n* sentences despite readers engaging in studious reading? Yes, there will always be memorisation freaks who pull off dazzling feats of recall. But is the common-or-garden reader swatting for an exam likely to find their study strategies thwarted by long paragraphs?

The American psychologist Walter Kintsch has conducted a number of experiments on paragraph recall. Unlike Horn, Kintsch well understood that sentence count alone is too blunt a measure when sentences can range widely in clausal complexity. Even clauses contain individual bits of information the complexity, or number, of which might affect recall. Rather than focus on sentence number or clause number, Kintsch decided to test whether conceptual density and conceptual uniqueness influences recall.[1] In one experiment, Kintsch gave subjects texts of varying conceptual densities and, after they had said they had had enough time to learn what was in the texts, they were asked to recall the propositions (or main points) made in each text. Subjects did not have to repeat verbatim what they had read; just recall, in their own words, what the salient points were in a text they had just read, and read without a time limit.

Kintsch found that subjects could recall less information from a paragraph that had more uniquely mentioned concepts than from a paragraph with fewer uniquely mentioned concepts *even if the paragraphs had the same number of sentences*. So *conceptual density*—the number of uniquely mentioned concepts—is more a determinant of successful recall than paragraph length. The greater the number of once-mentioned concepts, the more difficult it will be to accurately recall a text's main points, even when you have been given as much time as you like to study the text. Despite the repeated claim that Information Mapping is "research-based", there is no mention of Kintsch's research in Horn's book.

But what is more pertinent to our discussion are the *absolute* recall rates Kintsch discovered. Even with short paragraphs—those with just two sentences, and 21–23 words in total—recall rates as low as 58% were observed.[2] The recall rates for longer paragraphs—those with three sentences and 67–75 words in total—were lower (down to 36%), but that is not relevant to this discussion (and hardly surprising). What is relevant is that even with short paragraphs—those of just 2 sentences—recall rates were as low as 58%. So if there is a limit to paragraph length below which there is perfect recall—that is, a recall rate of 100%—that limit must be less than 2 sentences. Thus a close-to-perfect recall rate is only going to be possible with paragraphs of one sentence. In other words, if there is a chunking limit for perfect recall, it can only be one.

1. Kintsch W et al., "Comprehension and recall of text as a function of content variables", *Journal of Verbal Behavior and Verbal Learning*, vol. 14, iss. 2, 1975, pp. 196–214.
2. ibid., p. 202, table 7. The number of propositions—that is, amount of information—subjects were asked to recall, and the size of each paragraph, are given in table 1 on page 198. The size of the paragraphs is given in terms of the number of *words*. The sample texts Kintsch gives tells us the size of the paragraphs in terms of the number of *sentences*.

This would seem to contradict the claim we made at the end of the last section. When readers engage in studious reading they should be able to recall more than the contents of one sentence per paragraph. How else do we succeed at school and university? The answer lies in motivation: the more motivated we are the more likely we will adopt special learning strategies, such as summarising, rehearsal and mnemonic encoding. Kintsch's subjects were obviously not especially motivated to recall what they were asked to read.

It should be clear now that a chunking limit based solely on sentence-number is worthless as a guide to maximising recall. But might there not be an upper limit to sentence-number beyond which all recall drops to zero? If so, perhaps writers need to be take care not to breach this limit. But Kintsch's research suggests that it is not paragraph length that influences our ability to recall, but the number of uniquely mentioned concepts (or what we've called conceptual density). Thus a 15-sentence paragraph could, in principle, be easier to recall than a 7-sentence paragraph if it had fewer uniquely mentioned concepts. Perhaps there is an upper limit to the number of uniquely mentioned concepts in a paragraph beyond which recall fails entirely regardless of the motivation of the reader. Kintsch didn't explore this and the literature appears silent on the matter.

This is no great matter, since most of us do not read for recall. Instead, we engage in transient reading: reading without expectation of being asked to regurgitate the main points of what we have read.

Later research

We have just seen that paragraph length—understood as the number of sentences in it—is a poor indicator of the likelihood of correct *recall*. But might it be a better indicator of our ability to comprehend a paragraph when we are engaging in *transient* reading? We challenged this idea earlier, but let's assume, for the sake of argument, that this is still an open question.

How can we measure comprehension of material read transiently? Amount recalled would be a poor measure of transient understanding if those whose understanding is being measured are aware that it will be measured by the amount of material they can subsequently recall. It is only natural for subjects to want to get the best score possible on a comprehension test, and thus they will be disposed to adopt studious reading rather than transient reading as their approach. However, someone can understand what they are reading without necessarily being able to subsequently recall it. Thus transient understanding is best measured in some way other than by recall.

Perhaps aware of this Heisenberg-like uncertainty—that as soon as you try to measure *understanding* it becomes something else: *cognitive lodgement*, perhaps—many experimenters have opted instead for qualitative research (such as an analysis of readers' own assessments of ease of understanding).

In 1988, Cambridge University psychologist Heather Stark conducted an experiment in which subjects were asked to rate three texts on ease of reading, text coherence and text quality. Reading speed was also measured. The three texts were taken from Bertrand Russell (21 sentences over 4 paragraphs), George Orwell (54 sentences over 7 paragraphs) and Joan Didion (54 sentences over 6 paragraphs). The texts were presented to subjects in one of three formats: with the original paragraphs in place, with no paragraphs at all, and with the paragraph markers moved.

Stark's results will surprise many:

> "… the use or misuse of paragraph boundaries had no measurable effect on subjects' reading rate or ratings of ease of reading, coherence, or goodness … [It] doesn't seem to make a difference whether a text is explicitly divided into paragraphs or where the paragraph cues occur".[1]

In other words, the *unparagraphed* versions—even those that were 54 sentences long—were considered just as easy to read, of similar coherence and of similar quality to the paragraphed versions. Stark concludes her paper thus:

> "Given the persistent intuition that paragraph markings make text easier to read, it is surprising that the current study provided no support for this idea. Reading speed and ratings of ease, coherence, and goodness were not affected by the presence of or position of paragraph cues."[2]

An experiment reported in 1992 found similar results.[3] Three consecutive paragraphs were taken from a journal article on compact discs, likewise from a journal article on poverty, and likewise from a journal article on fluoridation (giving nine paragraphs in all). For each domain, the text in the three paragraphs was presented to subjects either as (a) the original three paragraphs (each containing about 80 words or 4 sentences), (b) merged into two paragraphs (each containing about 120 words or 6 sentences) or (c) merged into one paragraph (of about 240 words or 12 sentences). Subjects read the blocks of text and were asked to give a Likert response to a number of questions, one of which was "I found it easy to understand the writing". (A Likert response is

1. Stark, HA, "What do paragraph markings do?", *Discourse Processes*, vol. 11, 1988, p. 294.
2. ibid., p. 297.
3. Markel M, Vaccaro M & Hewett T, "Effects of paragraph length on attitudes toward technical writing", *Technical Communication*, Society for Technical Communication, vol. 39, iss. 3, 1992, pp. 454–6.

one from a set ranging from *I strongly agree* to *I strongly disagree*.) The conclusion:

> "… paragraph length did not affect the readers' attitudes towards the expertise of the writer, *the ease of comprehension*, or the quality of the passage … [Paragraph] length is not such a dominant textual feature that it affects … the ease of comprehension."[1]

These experiments did not test comprehension, but only readers' assessment of the ease of comprehension. In other words, they sought qualitative rather than quantitative results. Some subjectivity is perhaps unavoidable: some subjects could have been inclined to rate material as easy to understand even though they didn't understand it. Still, the two experiments gave remarkably similar results.

Another approach

There is a way to measure comprehension of transient reading while minimising the risk that readers will switch to studious reading during testing, namely, by using a cloze test. In a cloze test, subjects are given texts to read in which every fifth or sixth word has been deleted. They are asked to fill in the missing words. The number of correct words entered is a good measure of subject's comprehension of the text.

A cloze test can be used to test whether paragraph length—as measured by the number of sentences—affects comprehension. If Information Mapping—and folk wisdom—is correct, subjects faced with a text composed of a number of small paragraphs should score higher in a cloze test than subjects who are faced with the same information but presented in a single chunk of concatenated text. That is the hypothesis the experiment described below set out to test.

Materials

A snippet of text was selected that was neither too simple nor too technical. The snippet was taken from *What to do in an emergency*, published by Readers Digest in 1987. The snippet was prepared in two ways:

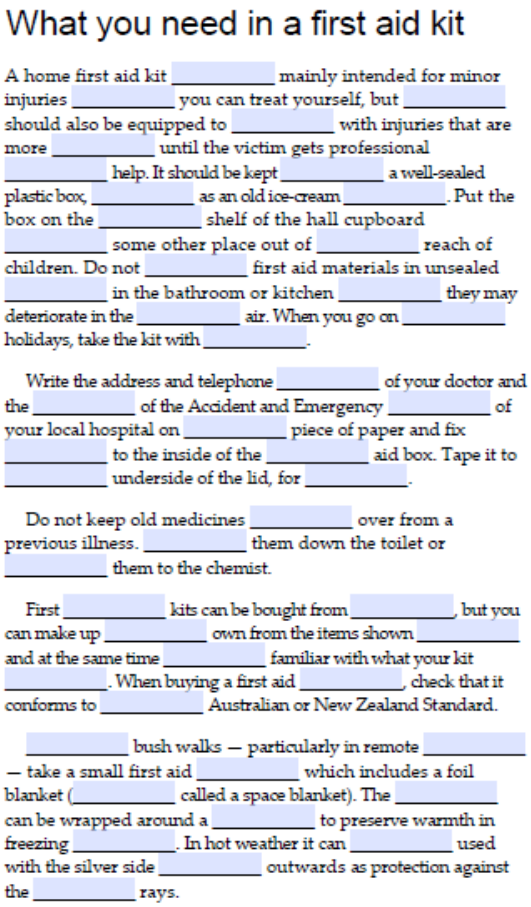
A extracted as-is: five distinct and consecutive paragraphs, comprising 15 sentences in total

B the same five paragraphs (and 15 sentences) but concatenated into a single block of text without standard paragraph indicators.

Snippet A was manipulated so as to make the start of each new paragraph obvious. Not only was the first line of each paragraph indented (as in traditional publishing), but additional space was added between paragraphs (in line with modern practice).

Every sixth word in both snippets was then removed (leaving 47 blanks in each snippet). Subjects would be asked to provide the missing words.

Both snippets were further prepared as a PDF form. This would enable subjects to respond online. <XREF>Figure 1 shows snippet A as it was presented to subjects. <XREF>Figure 2 shows snippet 2. Both snippets included a clickable button that enabled subjects to return the completed form via email.



What you need in a first aid kit

A home first aid kit _____ mainly intended for minor injuries _____ you can treat yourself, but _____ should also be equipped to _____ with injuries that are more _____ until the victim gets professional _____ help. It should be kept _____ a well-sealed plastic box, _____ as an old ice-cream _____. Put the box on the _____ shelf of the hall cupboard _____ some other place out of _____ reach of children. Do not _____ first aid materials in unsealed _____ in the bathroom or kitchen _____ they may deteriorate in the _____ air. When you go on _____ holidays, take the kit with _____.

Write the address and telephone _____ of your doctor and the _____ of the Accident and Emergency _____ of your local hospital on _____ piece of paper and fix _____ to the inside of the _____ aid box. Tape it to _____ underside of the lid, for _____.

Do not keep old medicines _____ over from a previous illness. _____ them down the toilet or _____ them to the chemist.

First _____ kits can be bought from _____, but you can make up _____ own from the items shown _____ and at the same time _____ familiar with what your kit _____. When buying a first aid _____, check that it conforms to _____ Australian or New Zealand Standard.

_____ bush walks — particularly in remote _____ — take a small first aid _____ which includes a foil blanket (_____ called a space blanket). The _____ can be wrapped around a _____ to preserve warmth in freezing _____. In hot weather it can _____ used with the silver side _____ outwards as protection against the _____ rays.

Figure 1: Snippet A, showing paragraph indicators

1. ibid., p. 455f. Emphasis added.

What you need in a first aid kit

A home first aid kit _____ mainly intended for minor injuries _____ you can treat yourself, but _____ should also be equipped to _____ with injuries that are more _____ until the victim gets professional _____ help. It should be kept _____ a well-sealed plastic box, _____ as an old ice-cream _____. Put the box on the _____ shelf of the hall cupboard _____ some other place out of _____ reach of children. Do not _____ first aid materials in unsealed _____ in the bathroom or kitchen _____ they may deteriorate in the _____ air. When you go on _____ holidays, take the kit with _____. Write the address and telephone _____ of your doctor and the _____ of the Accident and Emergency _____ of your local hospital on _____ piece of paper and fix _____ to the inside of the _____ aid box. Tape it to _____ underside of the lid, for _____. Do not keep old medicines _____ over from a previous illness. _____ them down the toilet or _____ them to the chemist. First _____ kits can be bought from _____, but you can make up _____ own from the items shown _____ and at the same time _____ familiar with what your kit _____. When buying a first aid _____, check that it conforms to _____ Australian or New Zealand Standard. _____ bush walks — particularly in remote _____ — take a small first aid _____ which includes a foil blanket (_____ called a space blanket). The _____ can be wrapped around a _____ to preserve warmth in freezing _____. In hot weather it can _____ used with the silver side _____ outwards as protection against the _____ rays.

Figure 2: Snippet B, showing no paragraph indicators

Method

The following were invited to participate in this experiment:

- students enrolled in *Technical Writing and Editing* at Melbourne University in 2011
- subscribers to *austechwriter* (an internet discussion forum for technical writers)
- participants in technical writing and scientific writing courses held by Abelard Consulting during 2011.

Participants were asked to fill in the missing words (and to leave blank any that were not immediately obvious to them).

When sufficient responses had been received, the number of correct words in each response was calculated. Close synonyms were accepted where the original word was not provided.

The average number of correct answers was then computed for each group of subjects. These averages (or means) were then compared using a *t*-test (for independent samples) in Stata statistical software.[1] All statistical tests were two-sided and a *p*-value < 0.05 was considered statistically significant.

1. *Stata statistical software*, release 11.0, StataCorp, StataCorp, College Station, 2010.

Results

	Responses	Mean	Standard Deviation
A: paragraphs	36	43.64	2.11
B: No paragraphs	58	44.67	1.79

There is no statistically significant difference between the two means (with $t = 2.54$, degrees of freedom = 92 and $p = 0.01$).[2] In other words, the very small difference between the means is just as likely to have occurred by chance as to have been caused by some cognitive mechanism or other.

This analysis assumes that the samples exhibit normality (that is, they fall within the typical bell-shaped distribution). The data did show some skewing, making it possibly better suited to a non-parametric analysis, such as the Wilcoxon rank sum test. That test gives much the same result: there is no statistically significant difference between the two distributions (with $z = 2.26$ and $p = 0.02$).

Conclusion

This experiment suggests that there is no loss of comprehension if as many as 15 sentences of material are presented to readers unchunked. This is nearly double the maximum paragraph length recommended by Information Mapping (and by Emeritus Professor Pam Peters, quoted on page 15). This is in line with our analysis of transient reading habits described in the previous part of this paper. Readers do not notice the size of the chunk they are reading or, if they do, it does not affect their understanding of it. Chunking, in other words, seems not to be necessary for comprehension.

So why chunk?

Chunking of informational text into paragraphs is not unimportant. There are many reasons why we do it and should continue doing it. For instance, paragraphing:

- "relieves the forbidding gloom of a solid page of text"[3])
- enables writers to present their ideas in logically related chunks
- meets readers' expectations that ideas are being presented in discrete and cohesive chunks

2. My thanks to Dr Gillian Dite, Centre for MEGA Epidemiology, University of Melbourne, for help in analysing the data.
3. Hudson, N, *Modern Australian Usage*, OUP, Melbourne, 1993., p. 294.

- satisfies the reading practice of the skimming reader (by enabling writers to present each of their main points at an easily identifiable place, namely, the first sentence of each paragraph)
- enables readers to quickly locate text (when paragraphing is combined with titling, or what Horn calls *labelling*)

The point of this paper has not been to discount the practice of paragraphing (or at least the practice of creating *well-formed* paragraphs). Rather, it has been to show that *comprehension* seems to be unrelated to paragraph length. We chunk to help readers engage with texts, to help them find the information they are after, to satisfy their expectations about the purpose of paragraphs, and to help them get the essence of what we are saying if they haven't time to read every word. But such chunking doesn't help readers *understand* what they are reading. That is what the research discussed above clearly suggests, namely, that unparagraphed text is considered by readers to be just as easy to read as paragraphed text.

To argue that comprehension is unrelated to paragraph length is not to imply that paragraphs can be of any length. Although the experiment conducted by Markel, Vacarro and Hewett (discussed on page 18) showed that subjects considered unparagraphed chunks of text as easy to comprehend as paragraphed chunks, it also showed that subjects *preferred* that the unparagraphed chunks—those composed of concatenated paragraphs—to have been paragraphed:

> "… regardless of which of the three passages they were reading, [subjects] felt that the 1-paragraph and 2-paragraph versions would benefit from shorter paragraphs, but that the 3-paragraph versions would not."[1]

This is hardly surprising given readers' expectations about the purpose of paragraphs. A reader who reads a chunk of text formed by the concatenation of three well-formed paragraphs is likely to sense the discontinuity between the merged topics. They will detect that the chunk is about three distinct topics and wonder why it has not been written according to the well-entrenched convention of one-topic-per-paragraph. So a preference for chunking-by-topic is to expected. We are habituated to read material chunked in that way.

If writers heed readers' expectation that a paragraph is the container for one topic and that all the sentences in it are related to that topic, then long paragraphs should be rare. Most writers simply don't have more than perhaps 10 things to say about any single indivisible topic. (And if the topic is logically divisible, then it should—if it is to meet readers' expectations—be split into its indivisible parts.) But the important point to note here is that paragraph chunking is better determined by *logic* than *length*. Readers expect *discreteness* and *cohesion* in a paragraph: *one* idea supported by a number of *related* sentences. If a writer happens to have, say, 10 things to say about one discrete idea, the paragraph will need to be ten sentences (or at least 10 clauses) long. To break that one paragraph into two solely on the grounds of the number of sentences in it rather than their logical connectedness is almost certain to distract the reader. The expectation that the second paragraph is presenting a new topic will not be met and the reader will likely be distracted.

To sum up: a chunking limit based on sentence number overlooks the fact that a paragraph is considered by readers to be a *logical* unit. To split a logical unit into two is just as distracting to readers as concatenating different logical units into the one paragraph. Readers' prime expectations about a paragraph—*discreteness* and *cohesion*—are not met. Whatever chunking limit we might apply to a paragraph should not be a fixed number or range (such as Horn's 7 ± 2). It should be a combination of two numbers: one fixed, one moveable. The fixed limit is the number of discrete topics presented in the paragraph. It should be fixed at 1. The moveable number is entirely dependent on what the author is attempting to convey. It is a number that defines the paragraph's degree of cohesion (or how well the sentences are related to a single topic). A sentence count in excess of this limit is what writers should be wary of, not a sentence count in excess of some fixed and seemingly artificial value, such as 7 ± 2. If a paragraph has eight sentences and only six are about the same topic, then the cohesion limit is six and the author has exceeded it by two. That is reason to take a scalpel to the paragraph. If a paragraph has 12 sentences and all 12 are related to the one topic, then the cohesion limit is 12 and the author has not exceeded it. It is fine as it is, and the author should have no fears that it has exceeded some comprehension limit.

In a word: a paragraph should be as short as possible but as long as necessary. It is as short as possible when it is about a single, indivisible topic; it is as long as necessary when everything that the writer wanted to say about that topic is contained within it. To insist on a sentence-number limit fails both logic and science.

1. op. cit., p. 455

To conclude

Over the course of three papers presently consecutively in this journal, we have presented proof that:

- Information Mapping—or at least its claim that cognition is limited to 7 ± 2 chunks—is not based on George Miller's research nor on any research that Miller quoted. To claim that it is ignores, or misunderstands, what Miller actually wrote and said.
- Even so, Miller's research has been superseded by more recent studies.
- The way Information Mapping applies the chunking principle is at odds with the way people read texts. The parent or sibling structure of a text is rarely if ever noticed by readers (whether seeking spur-of-the-moment information or in the exceptionally rare case where a reader reads the text all the way through).
- Even if readers did notice the parent or sibling structure of a text, no plausible evidence has been adduced to support the claim that the limited capacity of our *short-term* memory helps us comprehend anything other than atomic information (such as a clause or sentence). At the level of chapters, sections, maps and blocks, the capacity of our short-term memory appears entirely irrelevant.
- To understand a chunk is to understand the sentences in it and such understanding is unaffected by the number of sibling sentences that have to be read.
- Research in cognitive psychology—by Kintsch, Stark and others—shows that the number of sentences in a paragraph does not affect a reader's comprehension of that paragraph, contrary to the claims of Information Mapping and the recommendations found in many language handbooks.
- Chunking is valuable but is better based on chunk discreteness and cohesion than on chunk size. Logic, not length, is a better guide to quality paragraphing.

Geoffrey Marnell

geoffrey@abelard.com.au