Psychological Review
1994, Vol. 101, No. 2, 353-356

Copyright 1994 by the American Psychological Association, Inc.
0033-295X/94/$3.00

The Magical Number Seven: Still Magic After All These Years?

Alan Baddeley

The "magical number seven" gives a beautifully clear account of information theory and demonstrates how the concept of limited channel capacity can be applied across a range of sensory dimensions. However, its major influence stems not from this but from the demonstration that immediate memory span is relatively insensitive to amount of information per item. In emphasizing the importance of the *recoding* of information and developing the concept of chunking, Miller set the agenda for the next phase of cognitive psychology in which information-processing concepts went beyond the confines of information theory. This article continues to be cited because these underlying ideas continue to be fruitful.

As a long-time admirer of this article, I continue to quote it, but had not read it for some years, and began my preparations by rereading it while visiting the University of Otago in New Zealand. I was intrigued to discover that the 1956 volume of *Psychological Review* opened virtually automatically at this article, which in contrast to the pristine state of the rest of the journal was distinctly dog-eared, as if nibbled by generations of hungry kiwis. The obvious degree of use spoke well of the good taste of generations of Otago students, or at least of their teachers, and encouraged me to do the dog-earedness test on a couple of adjacent issues, coming up with Underwood's classic article on proactive inhibition (which I had in fact suggested as a possible candidate for this review of classics), and Hebb's article on the conceptual nervous system. On returning to Cambridge and attempting to find the article in the Applied Psychology Unit library, the 1956 *Psychological Review* proved to be the one 1950s issue that was out. A colleague was rereading George Miller's article because it is beginning to become influential among music theorists. There is, I think, little doubt that *The Magic Number Seven* is alive and well; but why?

The article operates at three separate levels. First of all, it offers a beautifully clear exposition of Claude Shannon's mathematical theory of information. It does so totally without recourse to mathematics and in terms that are immediately comprehensible to the novice. Second, it uses the device of the magic number seven as a basis for reviewing the application of information theory to absolute judgment. Third, it moves on to memory span, demonstrating the need to go beyond information measures, emphasizing the importance of recoding, and introducing the novel and important concept of *chunking.*

Information Theory

The concept of information is introduced and related to a range of more familiar concepts including both news value and variance, and its use is elegantly illustrated. Miller made it clear that the importance of information theory comes, first of all, from the general concept of the brain as an information-processing machine, a concept that has come to dominate cognitive psychology since that time. The idea of information as abstract, but nevertheless measurable, allows the theorist to draw conclusions, not only across different sensory and perceptual domains but even more widely, as Miller demonstrated by applying the concept of limited capacity to sensory judgments, tachistoscopic perception and, of course, memory span. The information-processing metaphor and the general utility of the concept of limited channel capacity have been enormously influential in the intervening years and continue to be valuable. This article, along with Broadbent's (1958) *Perception and Communication,* Miller, Galanter, and Pribram's (1960) *Plans and the Structure of Behavior,* and Neisser's (1967) *Cognitive Psychology,* played an important role in nurturing the fledgling discipline of cognitive psychology. As a framework for introducing information theory to his fellow psychologists, Miller adopted the ingenious stratagem of proposing that the limitation on information processing is set by "the magical number seven." As an expository device, I think this was brilliant, allowing Miller to link together a range of phenomena and generate what is arguably the best title in psychology, combining as it does the underlying concept of the general limit to cognitive processing capacity, with a tongue-in-cheek hint of mysticism and numerology.

Absolute Judgment

How convincing is the argument? In the case of absolute judgment, even for single dimensions, the number of categories that a subject can simultaneously handle averages about 6.5. The range is from 3 to 15 categories, prompting Miller to comment that "I find this to be a remarkably narrow range." Well, perhaps. As Miller pointed out, however, the number of absolute categories seems very small when one considers, for example, the number of faces of people we can recognize, prompting him to move on to considering multidimensional stimuli and, of course, the issue of what constitutes a single dimension, a topic that has been subsequently explored in considerably greater detail by Garner and his colleagues (Garner, 1962).

Another important factor in absolute judgments was subsequently pointed out by Chapanis and Overbey (1971), who were interested in the absolute judgment of colors. They went to considerable lengths to study the capacity of subjects to name spe-

Correspondence concerning this article should be addressed to Alan Baddeley, Medical Research Council Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF England.

cific colors and demonstrated that when subjects were given appropriate labels such as "pale bluish green," they were able to perform consistent and accurate absolute judgments on a set of 36 colors, with little or no need for practice, in contrast to the standard procedure in which, even after considerable practice, subjects were only able to manage to handle only a few more than seven different colors. This result points to the importance of response labeling, and the utilization of earlier experience in such absolute judgment tasks, and casts yet further doubt on the magicality of the number seven when applied to absolute judgments. I suspect, however, that this is not the feature of Miller's article that principally accounts for its continued influence. For this, one must turn to Miller's application of the concepts of information theory to the analysis of memory span.

Memory Span

Miller ingeniously linked the limit on absolute judgment to memory span by suggesting that the sequential presentation of items provides a way of circumventing the limited capacity for absolute judgment, in short, that one considers "memory as the handmaiden of discrimination." At this point, note that Miller explicitly warned the reader against being seduced by the fact that memory span is typically about seven items:

> I have just shown you that there is a span of absolute judgment that can distinguish about seven categories and that there is a span of attention that will encompass about 6 objects at a glance. What is more natural than to think that all three of these spans are different aspects of a single underlying process? And that is a fundamental mistake as I shall be at some pains to demonstrate. (Miller, 1956, page 91)

He went on to demonstrate the crucial difference between the limitations on span and on absolute judgment, with judgment being limited by the amount of *information,* measurable in bits, whereas immediate memory span is determined by the *number* of items, or to be more accurate, the number of *chunks.* Here he introduced the concept that lies at the heart of the article, namely, the recoding of incoming information, concluding that "the process of memorization may be simply the formation of chunks, or groups of items that go together, until there are few enough chunks so that we can recall all the items" (Miller, 1956, p. 94), a conclusion that still commends itself to many theorists and does, of course, form an important component of Allen Newell's unified theory of cognition, Soar (Newell, 1990).

What Happened to Information Theory?

Although Miller's article presents a beautifully clear exposition of information theory, I suspect that this is not what has encouraged people to continue reading and citing it. In the late 1950s and early 1960s, information theory seemed likely to transform experimental psychology and form an essential component of any psychologist's education, and yet it is now rarely mentioned. Why should that be?

The information-processing approach taken by Miller had two components; the first involved the general concept of the organism as an information-processing system, whereas the second comprised a specific mathematical theory of information that allowed the capacity of the system to be accurately measured. Although the information-processing metaphor has been enormously influential in developing the field that became known as cognitive psychology, the precise measures of information-processing capacity have proved to be much less valuable.

The problems of applying information theory to psychology show up particularly clearly in the study of reaction time, in which Hick's Law and Fitts's Law initially seemed to offer some elegant truths about the human operator. Hick (1952) observed that choice reaction time increased linearly with the log number of choices, supporting the concept of the human operator as an information-processing channel of limited and measurable capacity. A similar conclusion seemed to follow from the elegant demonstration by Fitts (1954) of a lawful relationship between the rate of tapping two adjacent targets and their size and intertarget distance. Both laws reflect the limited channel capacity of the human operator, as measured by using Shannon's mathematical theory of information. However, it rapidly became obvious that the human operator is not like a static electronic device that has a fixed and immutable information-processing capacity. Mowbray and Rhoades (1959) took full advantage of a captive subject panel provided by the local prison to explore the influence of many hours of practice on Hick's Law. They observed that the more their subjects practiced, the flatter became the function relating reaction time to number of alternatives, suggesting that the system was changing so as to be capable of processing ever-increasing amounts of information. Eventually, the slope virtually disappeared; did that therefore suggest that the rate of information processing was infinite?

Other studies suggested that one did not even need excessive amounts of practice, provided the compatibility between stimulus and response was great enough. In one study, Leonard (1959) used vibrators attached to the subjects' fingers as the stimuli, with the pressing of the appropriate digit as the required response. Once again, the slope reduced to virtually zero, a result that was also obtained by Davis, Moray, and Treisman (1961), using the repetition of auditorily presented items as their response and finding no increase in latency as set size increased, thus anticipating the later work of McLeod and Posner (1984) that indicated a "privileged loop" between hearing and echoing back spoken items.

Hence, the initial hypothesis that information theory offered a royal road to the analysis of skill proved illusory. However, the deviations from the original informational model laid the foundations for important new concepts ranging from S–R compatibility (Fitts & Switzer, 1962), through automaticity (Schneider & Shiffrin, 1977), to current concerns with the role of attention in the control of action (Baddeley, 1993; Shallice & Burgess, 1993). Indeed, it could be argued that the central issue facing attempts to model the attentional control of action continues to be that of how the organism uses learning to facilitate skilled performance while at the same time allowing habits to be overridden when the need arises.

In the area of language, I would argue that information theory was more immediately successful because it emphasized the redundancy of language and led to a much more detailed investigation of linguistic structure. The classic earlier article by Miller and Selfridge (1950) had used the generation of approximations to English prose as a means of demonstrating the dependence of memory on the information contained in text. Tulving and Patkau (1962) took this work one stage further, demonstrating

that if one defined a chunk as being a sequence of words recalled by the subject in the order presented, then as the passages approximated more and more closely to English, the total number of words systematically increased while the number of chunks remained constant, a very nice application of the concept of chunking developed by Miller in the "magic number" article.

During the late 1950s and early 1960s, I myself was concerned with the practical problem of attempting to design postal codes that were readily memorable and rapidly typeable and found information theory to be useful, both in generating memorable codes and in predicting the memorability of existing codes. In the case of traditional CVC nonsense syllables, I found that a measure of predictability, based on applying the statistical letter structure of English to the constituent items, was a consistently better predictor of memory than the ratings of association value that were used standardly at the time (Baddeley, 1963). I even generated memorable postcodes for every town in Britain but, alas, the Post Office had other ideas!

However, in the study of language, the analysis of syntax became the dominant theme, spearheaded by Chomsky in alliance with George Miller and their Project Gramarama (Miller, 1962), which in turn led Miller on to a concern with the deeper issues of semantics (Miller & Johnson-Laird, 1976). At the level of the word and the letter, however, I think information theory still has a role to play. Furthermore, because of its written form, we know a great deal more about the statistical structure of language than we know about other aspects of cognition and behavior, which may be equally sensitive to the statistical structure of the environment. As our technical capability for measuring the statistical structure of the physical environment increases, we are beginning to see a revival of measures and concepts based on information theory in the analysis of sensory processing (Atick, Li, & Redlich, 1992).

However, as Miller demonstrated so astutely in the article we are celebrating, a major challenge to a simple channel capacity interpretation of cognition comes from the propensity of human subjects to recode information. Because the nature and extent of such recoding is typically dependent on previous learning, a variable on which people can vary enormously, the prospect of coming up with a single quantitative measure of processing capacity becomes increasingly remote. The situation is further complicated by the possibility of setting up hierarchical structures of chunks. If seven chunks can be held, can each one be divided into seven subchunks? Presumably not, because that would suggest that one can hold 49 chunks. Perhaps the number seven, itself, comes from chunking; Broadbent (1971) for example, suggested a capacity of three, with each chunk perhaps able to hold three further chunks (the magic number 7 + 2?). Mandler (1967), on the other hand, opted for five as the magic number (7 − 2?). My own view is that it is unlikely that the limit is set purely by the number of chunks, independent of such factors as the degree to which material within each chunk is integrated as a result, for example, of prior learning. The relationship between chunks may also vary. In a narrative passage, there may be very strong constraints that are likely to make such a passage easier to recall than a purely descriptive passage of equivalent length (Bartlett, 1932). Thus, I think Miller was correct in describing recoding as "the lifeblood of thought processes," and to emphasize the importance of this rather than amount of information or, indeed, number of chunks.

What Happened to Immediate Memory?

In the 35 years since the publication of Miller's article, the study of immediate memory has drifted in and out of fashion. The concept of a limited capacity in terms of chunks has continued to feature in the textbooks but has not tended to play a particularly important theoretical role, with the notable exception of Herbert Simon's periodic contributions to the topic, which tend to be principally concerned with the important issue of what constitutes a chunk (Simon, 1974; Zhang & Simon, 1985). My own work, for example, could be regarded as focusing on variables that explicitly *change* the number of chunks that can be held in immediate memory. Phonological similarity, for instance, has a major impact on immediate memory span, which is dependent on the similarity between items rather than number of chunks (Baddeley, 1966; Conrad & Hull, 1964).

Another exception to the constant chunk hypothesis would appear to be provided by the influence of word length on immediate memory span, which is found to be linearly related to the spoken duration of the constituent words (Baddeley, Thomson, & Buchanan, 1975). As the words are unrelated, one might expect each word to constitute a chunk. The fact that span is strongly influenced by the spoken duration of the words suggests a system that is time based rather than chunk based. The concept of a phonological loop involving a time-based store and an articulatory rehearsal process that operates in real time offers a simple account of this and other related findings (Baddeley et al., 1975). The fact that the prevention of rehearsal by articulatory suppression removes the word length effect is also consistent with the phonological loop model while not being readily explicable in terms of the chunking hypothesis.

Zhang and Simon (1985) explicitly tackled the relationship between word length and chunking by using Chinese, a language that is ideally suited to experiments attempting to separate visual, auditory, and chunking factors in immediate memory. In one study, they tested their subjects' memory span for three types of visually presented ideographs. The three types of ideograph were all familiar and could reasonably be regarded as comprising ready-made chunks. They did, however, differ in spoken length, involving monosyllabic names, words comprising two syllables, and idioms involving four syllables. Mean recall decreased with syllabic length, resulting in spans of 6.6, 4.6, and 3.0, respectively. Span was clearly not a simple function of number of chunks. However, span measured in syllables was not constant either, with spans being 6.6, 9.2, and 12.0 syllables, respectively. Zhang and Simon concluded that there is a need to assume effects of both the spoken duration of the items, as proposed by the Baddeley and Hitch (1974) working memory model, and also of number of chunks. I accept this and suggest that the chunking effects may be dependent on the operation of the central executive component of working memory.

To the best of my knowledge, there has been relatively little recent application of the concept of chunking to visual memory. I suspect, however, that it could be applied with some success to a task such as that devised by Wilson, Scott, and Power (1987) involving the immediate memory for cells within a matrix, which I suspect subjects tend to recall in terms of locally organized chunks, a simpler form of the chunking that is observed when chess players are required to remember a game position (De Groot, 1965).

As mentioned earlier, any complete theory of immediate recall will need to cope with chunking as an important variable; I suspect that the number of chunks that can be maintained is limited in part by the capacity of the central executive of working memory, and that chunking may be based on many different factors, ranging from minute temporal pauses that can introduce prosodic factors into immediate auditory memory (Frankish, 1985) through to the importance of long-term memory in the chunking of complex material such as prose passages (Tulving & Patkau, 1962) and chess positions (De Groot, 1965).

Conclusion

George Miller has eloquently advocated the importance of "giving psychology away," presenting it sufficiently clearly and cogently that it is accessible to the nonspecialist and the layman. The "magic number seven" is a superb example of presenting a highly technical subject elegantly and simply. However, the reason that this article continues to be influential at a time when information theory is largely ignored within psychology stems from the insights that allowed Miller to go beyond the restrictions of the theory itself. In emphasizing the importance of recoding, Miller pointed the way ahead for the information-processing approach to cognition, and in developing the concept of chunking, he provided a concept that continues to be fruitful in the analysis of learning and memory. The article, if not the number seven, retains its magic.

References

Atick, J. J., Li, Z., & Redlich, A. N. (1992). Understanding retinal color coding from first principles. *Neural Computation, 4,* 559–572.

Baddeley, A. D. (1963). *The coding of information.* Unpublished doctoral dissertation, University of Cambridge, Cambridge, England.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology, 18,* 362–365.

Baddeley, A. D. (1993). Working memory or working attention? In A. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness and control. A tribute to Donald Broadbent* (pp. 152–170). London: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). San Diego, CA: Academic Press.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 575–589.

Bartlett, F. C. (1932). *Remembering.* Cambridge, England: Cambridge University Press.

Broadbent, D. E. (1958). *Perception and communication.* Elmsford, NY: Pergamon Press.

Broadbent, D. E. (1971). The magic number seven after fifteen years. In A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory* (pp. 2–18). New York: Wiley.

Chapanis, A., & Overbey, C. M. (1971). Absolute judgments of colors using natural color names. *Perception and Psychophysics, 9,* 356–360.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55,* 429–432.

Davis, R., Moray, N., & Treisman, A. (1961). Imitative responses and the rate of gain of information. *Quarterly Journal of Experimental Psychology, 13,* 78–90.

De Groot, A. D. (1965). *Thought and choice in chess.* New York: Basic Books.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47,* 381–391.

Fitts, P. M., & Switzer, G. (1962). Cognitive aspects of information processing: I. The familiarity of S–R sets and subsets. *Journal of Experimental Psychology, 63,* 321–329.

Frankish, C. (1985). Modality-specific grouping effects in short-term memory. *Journal of Memory and Language, 24,* 200–209.

Garner, W. R. (1962). *Uncertainty and structure as psychological concepts.* New York: Wiley.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4,* 11–26.

Leonard, J. A. (1959). Tactual choice reactions: I. *Quarterly Journal of Experimental Psychology, 11,* 76–83.

Mandler, G. (1967). Organization in memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 1, pp. 327–372). San Diego, CA: Academic Press.

McLeod, P., & Posner, M. I. (1984). Privileged loops from percept to act. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance* (Vol. 10, pp. 55–66). Hillsdale, NJ: Erlbaum.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Miller, G. A. (1962). Some psychological studies of grammar. *American Psychologist, 17,* 748–762.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior.* New York: Holt, Rinehart & Winston.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception.* Cambridge, England: Cambridge University Press.

Miller, G. A., & Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *American Journal of Psychology, 63,* 176–185.

Mowbray, G. H., & Rhoades, M. V. (1959). On the reduction of choice reaction times with practice. *Quarterly Journal of Experimental Psychology, 11,* 16–23.

Neisser, U. (1967). *Cognitive psychology.* New York: Appleton-Century-Crofts.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic information processing: I. Detection, search and attention. *Psychological Review, 84,* 1–66.

Shallice, T., & Burgess, P. (1993). Supervisory control of action and thought selection. In A. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness and control. A tribute to Donald Broadbent* (pp. 171–187). London: Oxford University Press.

Simon, H. A. (1974). How big is a chunk? *Science, 183,* 482–488.

Tulving, E., & Patkau, J. E. (1962). Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. *Canadian Journal of Psychology, 16,* 83–95.

Wilson, J. T. L., Scott, J. H., & Power, K. G. (1987). Developmental differences in the span of visual memory for pattern. *British Journal of Developmental Psychology, 5,* 249–255.

Zhang, G., & Simon, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory & Cognition, 13,* 193–201.

Received August 15, 1993
Revision received September 20, 1993
Accepted September 20, 1993 ■