# Crucial Errors in Miller's (1956) Application of Information Metrics in a Comparison of the Spans of Absolute Judgment and Immediate Memory[1]

Bruce L. Bachelder[2]
Independent Practice of Professional Psychology, Retired
Morganton, NC

Abstract

The "magical numbers" (Miller, 1956) are important to theory and still need to be explained (Shiffrin & Nosofsky, 1994, p. 360). When performance in the three tasks, the spans of absolute judgment, immediate memory, and apprehension, is plotted as probability of a perfect response against set size the curves are virtually identical inverse ogives whose inflection points correspond to the magical numbers, 7+/-2 (Bachelder, 2001). These data bolster unitary theories of the magical numbers (e.g., Bachelder, 2001; Bachelder & Denny, 1977a, 1977b; Cowan, 1999, 2001; Killeen & Taylor, 2000). Baddeley (1994, p. 354) challenged unitary theories with his assertion that Miller demonstrated "the crucial difference" between absolute judgment and immediate memory, limited by amount of information and number of chunks, respectively.

Miller's demonstration is flawed. He applied the $I_t$ method (Garner & Hake, 1951) to absolute judgment but the "weighted responses" method to immediate memory. The two methods yield very different estimates of information transmission, about 2.3 and 19.6 bits in absolute judgment versus immediate memory, respectively. When immediate memory is assessed via the $I_t$ method the "crucial difference" disappears. Immediate memory data scored via the $I_t$ method yield a channel capacity effect at about 2.5 bits of transmitted information, essentially identical to that observed in absolute judgment tasks.

Miller's errors obfuscated the fundamental similarity of the three tasks. The present analyses help remove one of the barriers to careful consideration of unitary theories of the magical numbers.

George A. Miller (1956), in his well-known "magical number 7" paper, brought the mathematics of information theory to bear on the *spans question*. The spans question asks, *What is the nature of the three curious span limits?* The nature of the span limits is one of the oldest topics in scientific psychology (Bachelder, 2005) and has been addressed by some of our most prominent early figures: Oliver Wendell Holmes (1871, p. 33, as cited by Dempster, 1981, p. 66), Sir William Hamilton (1836-1837/1880, pp. 176-177), William James (1890/1950, pp. 405-409; 676), Hermann Ebbinghaus (1885/1913, p. 47), James McKeen Cattell (1890), Alfred Binet (Binet & Henri, 1895, as cited by Dempster, 1981, p. 63; Binet & Simon, 1905/1916, pp. 53-54, 58-59, 61), and Wilhelm Wundt ("scope of attention"; 1896/1897, p. 211-213; 1911/1912, p. 24-

[1] This paper was the basis for a presentation at the 38th Annual Conference of the Society for Mathematical Psychology, Memphis, TN, August 7, 2005.

[2] Address communications to 306 West Union Street, Morganton, NC 28655-3729 or Bruce@BruceBachelderPhD.com web page: www.BruceBachelderPhD.com

25). More recently, Arthur R. Jensen (1964, 1970) emphasized the strong link with intelligence and Shiffrin & Nosofsky (1994, p. 360) stated, "the existence of the limitations and the importance of such data for theory remain unchallenged."

The traditional *multiple spans* answer to the spans question has it that the three limits reflect three different underlying limits. The *unitary span* answer has it that all three limits are just variations on a single limitation. Miller (1956, p. 92) concluded that the span of immediate memory and the span of absolute judgment are two different kinds of limits. I will argue today that Miller's conclusion was based on errors in his application of the mathematics of information theory. When those errors are corrected his analyses are entirely consistent with, even suggestive of a unitary interpretation of the spans.

Some might be tempted to dismiss my topic as a curious, even quaint concern with errors made over 50 years ago in the application of a mathematical approach no longer considered useful (cf. Sperling, 1988, p. 76); but Miller's paper isn't just any paper. It is the single most important paper on the spans question and it is one of the five documents which launched the cognitive revolution (Gardner, 1985). It is probably the most famous and most cited psychological paper of all time and it continues to be cited frequently to this very day. It was required reading in my graduate school days and my current working copy is a worn Xerographic copy of the Bobbs-Merrill reprint I used in graduate school in the late '60s.

The impact of Miller's paper has been huge. If it contains significant errors the potential negative impact is also huge. Errors in Miller's paper may have given rise to or perpetuated fundamental problems in the working paradigmatic presumptions of cognitive science. The multiple-spans interpretation is deeply embedded in cognitive science and prominent investigators have stated in a prominent journal that Miller's paper bolsters the multiple spans point of view: "[Miller demonstrated] the crucial difference between the limitations on span and on absolute judgment" (Baddeley, 1994, p. 354). In this case, an assertion of a "crucial difference" is logically equivalent to an assertion of the multiple-spans point of view. Shiffrin & Nosofsky (1994, p. 360) refer to "Miller's (1956) convincing and amusing arguments" against a deep connection among the three span limits. A "deep connection," of course, would be a unitary interpretation of the three span limits.

At a personal level, Miller's multiple-spans interpretation pits my unitary *span theory* (Bachelder, 1970/1971, 1977, 2001; Bachelder & Denny, 1977a, 1977b; Denny & Davis, 1982) against the magical number paper and George A. Miller. This is an uncomfortable situation I sometimes liken to the clash of David and Goliath.

Here's my plan for today:
1) I'll review the prototypical data of the three span tasks so you have the basics in mind.
2) I'll briefly introduce some of the reasons I think the spans question is important.
3) I'll briefly present some of the evidence which suggests the spans are unitary.
4) I'll detail Miller's analyses of the spans of absolute judgment and immediate memory. I don't have time for the span of apprehension.
5) I'll point out the errors which led to his conclusion that the span limits are of different natures and which led to his introduction of the concept of the chunk.

6) I'll correct those errors and show that his own analyses, correctly done, are actually consistent with a unitary limitation.
7) I'll close with a brief overview and discussion.

***The Three Span Tasks***

Figure 1 presents the prototypical data of the three span tasks in the way they were commonly presented before information theory became a hot topic. The figure plots probability of a correct response as a function of set size. The span limit is assessed as the 50% threshold of these curves. The span phenomena are highly robust. These data are taken from published papers spanning about 57 years. They derive from different laboratories and different paradigmatic points of view (schools of thought).

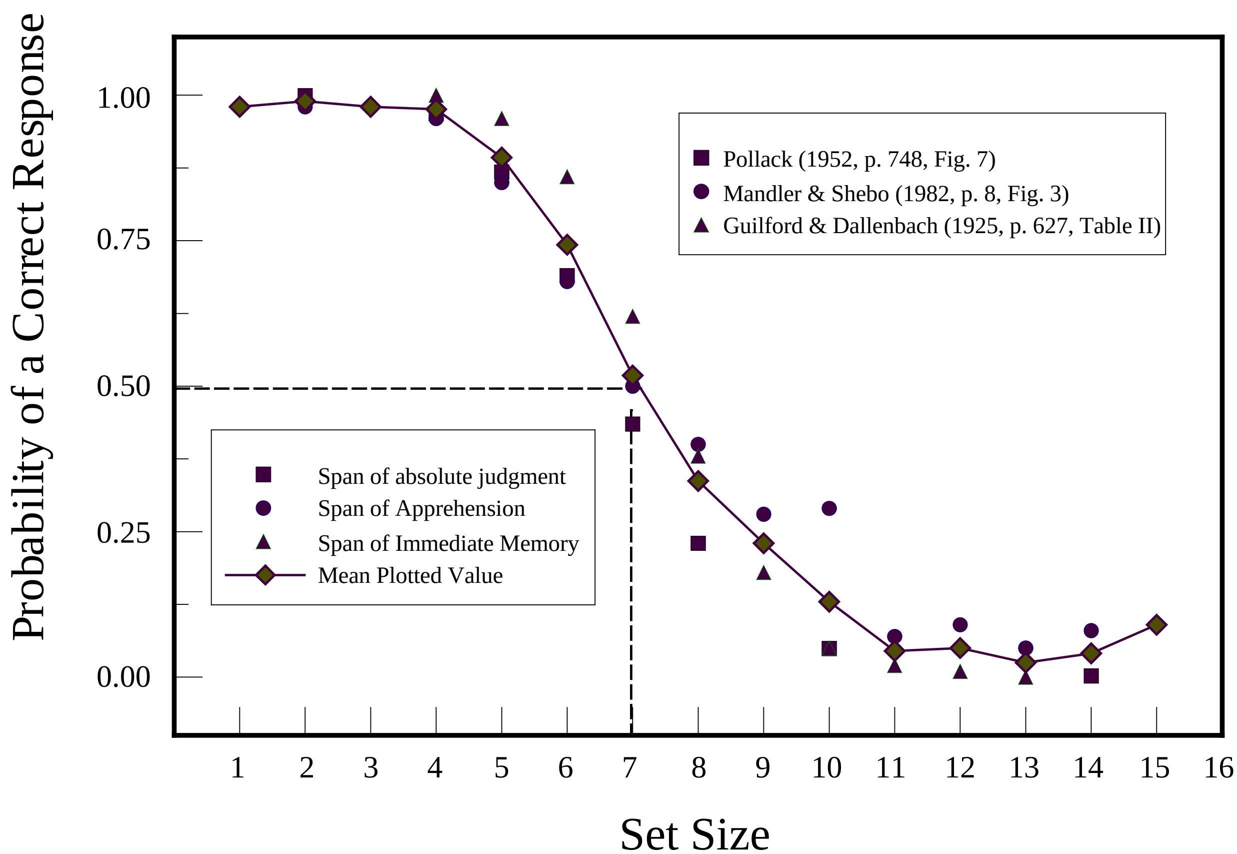


Figure 1. Probability of a correct full response as a function of set size. See Appendix for details of the presentation.

In the immediate memory span test subjects see or hear sets of randomly generated digits or words. College students repeat the sets reliably and perfectly only up to about 4 items. Thereafter, perfect performance drops sharply to near 0. The 50% threshold of this curve falls right around 7 items, that is, Miller's magical number.

The second span test, often called *span of apprehension*, but also *span of attention* and, less commonly, *span of numerosity*, has subjects see sets of randomly arranged items briefly presented (about .5 sec). College students report the numerosity reliably and perfectly up to about 4 items. Thereafter, perfect performance drops sharply to near 0. The 50% threshold of this curve falls right around 7 items.

Finally, *span of absolute judgment*, also called *span of absolute identification*, has subjects see or hear a single stimulus randomly selected from a set of stimuli varying along a quantitative dimension (for example, 3 squares varying in size or 5 tones varying in frequency). College students report relative magnitudes reliably and perfectly only up to sets of about 4 items. Thereafter, perfect performance drops sharply to 0 or near 0. The 50% threshold of this curve falls right around 7 items.

These curves might surprise those familiar with the span limits because the data for the three tasks are not usually presented in fully comparable ways. See Appendix for the details of this particular comparison.

***The Importance of the Spans Question***

***The potential to provoke scientific revolution.*** I have come to believe the spans question has the potential to move scientific psychology along from its current status as a fractionated preparadigmatic science toward status as a more unified paradigmatic science in the manner described by Thomas S. Kuhn (1970) in his book, *The Structure of Scientific Revolutions*. This move toward unification will occur in two broad ways. First, by developing a unitary conception of the three span limits, the previously distinct subdisciplines associated with the limitations on “immediate memory,” “apprehension,” and “absolute identification” will be merged. Second, the two major traditions, cognitivism and behaviorism, will be bridged and merged. Cognitive tasks, but not mentalism will be brought to a behavioral approach. The scientific rigor of behaviorism but not the tight restrictions on subject matter and approach will be brought to cognitivism. The result will be a point of view firmly rooted in both traditions. It will have the strengths of both and will side step major weaknesses of both.

The traditional interpretations of the three span limits is that they are limitations on memory, attention, and psychophysical processes, for the spans of immediate memory, apprehension, and absolute judgment, respectively. However, in the literature at large there is anything but a consensus on these presumptions. For example, it is almost a truism of cognitive science that the span limit in serial recall derives from an underlying limitation in memory, but one of the oldest and one of the newest theoretical interpretations of the immediate memory span view it as a limitation on attention, not memory; for example, Jacobs (1887), Wundt (1896/1997, pp. /211-123; 1911/1912, pp. 24-25), and Cowan, (1988, 1999, 2001). Woodworth (1938, p. 684) discussed the span of apprehension in his chapter on attention, but pointed out that the “span of attention” might just as well be called the “perception of number.” In span theory this limitation is called *Span of Numerosity*. Behaviorists have interpreted the memory span limit to reflect a limit on learning, conceived as an underlying process (Hilgard, 1951, p. 547), or to individual differences in repertoire deriving from differential learning history (Staats & Staats, 1964, p. 176). Peterson (1963, p. 351) suggested that viewing “memory” span in terms of learning would be fruitful. Bachelder (1970/1971) pointed out that “memory” span can also be viewed as “elicitation” span. Span of absolute judgment has long been considered a limitation in psychophysical process, but Miller (1956) viewed it as a limitation in channel capacity. In span theory I consider it to be an example of a pervasive and unitary limitation of the ability to cope with *span load*, a fundamental characteristic of tasks.

Clashing alternative points of view are the basis of one type of scientific discovery, the holy grail of science. They are also the basis for scientific revolutions. Since the clashing points of view on the spans question derive from just about all the schools of thought including early cognitivism, modern cognitive science, and several flavors of behaviorism, the ultimate answer to the spans question has the potential to unify the entire field, even cognitivism and behaviorism. I don't have time today to detail these fascinating matters, but I am pursuing them in other manuscripts.

***The Challenge to a Fundamental Paradigmatic Presumption of Cognitive Science.*** The main modern theoretical models of the three span tasks are built on the multiple-spans presumption; that the three spans are of fundamentally different natures. A unitary theory of the three limits challenges multiple-spans interpretations. Because of the natural inertia of science and scientists, such traditional working presumptions are difficult to change in a discipline at large even if the presumptions clearly conflict with valid data (Kuhn, 1970, p. 77).

Sometimes the very existence of these different models is suggested to be evidence against a unitary theory, but the current models were created based upon the multiple-spans *presumption*. The evidence for a unitary point of view is substantial and I am confident that a clever cognitive science modeler will find it straightforward to model the three span limits in terms of a unitary limitation. Cowan's attention interpretation is an excellent point of departure as is my own notion of *span ability*, the ability to cope with *span load*. Span load is a characteristic of tasks and among the three span tasks it is essentially equivalent to size of the stimulus set. In span theory the stimulus set is defined as the set of stimuli which are conjunctively relevant for correct target responding. Put in other terms, it is the set of stimuli which are individually necessary and collectively sufficient to specify a correct target response. The ogival function between performance and set size is what the three span tasks have in common. It is the link among the three tasks and the basis for a unitary understanding of the limitations.

***The Potential to Improve Our Understanding of Intelligence.*** I am just one of the more recent in a long line of thinkers who have believed that the span limits are implicated at a fundamental level in an understanding of human intelligence. Arthur R. Jensen (1964, p. 20) stated, "A thorough analytical understanding of the nature of memory span would probably constitute a major step toward understanding the psychological conglomeration called intelligence." In another paper he states, "The relationship of memory span to general intelligence is actually greater than is generally believed (Jensen, 1970, pp. 71-74)."

The idea that the memory span limitation lies at the core of intelligent function is as old as scientific psychology. Oliver Wendell Holmes, a philosopher, speculated in 1871 (p. 33, as cited by Dempster, 1981, p. 66) that the memory span test might prove to be useful as a "mental dynamometer," that is, as a measure of the power of the mind. William James (1890/1950, p. 405) stated "the number of *things* we may attend to . . . [depends on, among other things] the power of the individual intellect." Wilhelm Wundt (1911/1912, p. 42) considered the two processes, *apperception*, which he indexed by a span test, and *apprehension*, together to ". . . form the whole of our psychical life." Blankenship (1938, p. 17) reviewed the memory span literature and included several papers which put "memory span ability at the base of all intellection."

The span limits bear a striking relation to measured intelligence and status as mentally retarded or normal. My own interest in the memory span test came about because of its surprising ability to discriminate the mentally retarded from normal children and adults and levels of retardation among people with mental retardation. "Except in cases of special defects or organic disease, adults who *cannot* retain 5 digits forward and 3 backward will be found, in 9 cases out of 10, to be feeble minded or mentally disturbed (Wechsler, 1958, p. 71)." The memory span test was incorporated by Alfred Binet in the very first intelligence test (Binet & Henri, 1895, as cited by Dempster, 1981, p. 63; Binet & Simon, 1905/1916, pp. 53-54, 58-59, 61) and has been in all major Binet and Wechsler intelligence tests since then. Bachelder (1970/1971) found a correlation of 0.79 between digit span and IQ in a group of 274 retarded and borderline institutional residents. Horn (1968, p. 249, Table 1) cited a factor analysis which indicates the memory span test loads 0.50 on *fluid intelligence*, but 0.00 on *crystallized intelligence*. Wechsler's Digit Span subtest loads 0.63 (0.80, corrected for attenuation) with the *g* factor in Wechsler's (1958, p. 122) factor analysis of WAIS IQ scores for the age group 18-19 (as calculated by Jensen, 1970, p. 72). As a measure of intelligence, a "memory" span test has poor *face validity*, but excellent *construct validity*, arguably surpassing that of an IQ test. For a discussion bearing on this point see Bachelder & Denny (1977a, p. 130).

***The Argument for a Unitary Span.*** Even though the multiple-spans presumption is traditional and the three span tasks tend to look very different, there is ample reason to entertain a unitary hypothesis. There are strong procedural and quantitative similarities among the tasks:

1) Each task can be characterized by the *stimulus pool* from which test stimuli are chosen. For example, in digit span the stimulus pool is often the 10 digits and individual trials use a subset of them. Size of stimulus pool is a weak variable, at best, in memory span (Hayes, 1952). It can be argued that it is weak, at best, in all three tasks.

2) Each task can be characterized in terms of its *relevant stimulus set*, the set of stimuli necessary for correct responding on each trial. Set size, the number of items in a relevant stimulus set, is a potent variable in all three tasks and is the dimension along which the span limitations occur (Figure 1). Set size corresponds to *span load* in span theory.

3) When performance is assessed as probability of a correct response and plotted as a function of set size, the results are curves resembling inverse ogives which are virtually identical for all three tasks (Figure 1).

4) The three curves show a pronounced threshold effect which defines the magical number limitation in all three tasks. Generally, the absolute magnitudes of the span limits in all three tasks are highly similar when consistently measured.

5) Errors in both memory span and absolute judgment are distributed unevenly with errors tending to occur more for stimuli in the middle of the stimulus string in memory span and for stimuli in the middle of the stimulus range in absolute judgment. This effect is called the serial position effect in serial recall and the bow, edge, or end effect in absolute judgment.

6) As we will see later, both memory span and absolute judgment show the channel capacity effect Miller reported for absolute judgment.

7) There is solid reason from the subdisciplines of psychometrics and intelligence to expect the three tasks to be correlated because each task measures a *mental ability* as defined by Arthur R. Jensen (1986, p. 303)[3]. "All varieties of mental ability tests are positively correlated with one another in the general population" (Jensen, 1986, p. 304). This phenomenon is called *positive manifold* and Jensen considers it to have the status of a law of nature (Jensen, 1986, p. 304).

8) The span of immediate memory and $I_t$ in absolute judgment correlate high enough to suggest they both measure the same latent variable. Bachelder (1976, as cited by Bachelder & Denny, 1977a, p. 140) found a correlation of 0.78, $p < 0.0005$, (N=60) in a group of retarded, average, and bright adults.

9) All three spans covary with status as mentally retarded or nonretarded. Among the mildly retarded the "magical numbers" fall in the 5 +/- 2 range, rather than the 7 +/- 2 range of college students (Spitz, 1973).

10) The span limits in all three tasks are surprisingly insensitive to practice and training effects. For memory span see Bachelder & Denny (1977a, p. 30) and Blankenship (1938, pp. 11-12). Shiffrin & Nosofsky (1994, p. 357) mention the absence of practice effects for absolute judgment. Garner (1962, p. 77) mentions that it is normally presumed that learning plays no important part in determining channel capacity in absolute judgment. Research designs in both absolute judgment and span of apprehension indicate a presumption of no practice effects; usually no provisions are made to eliminate or counterbalance them. In absolute judgment (e.g. Garner, 1953, p. 374) large numbers of trials are given and early trials are discarded, reflecting that experience is needed to stabilize performance, but beyond some point the span (channel capacity) doesn't change with practice. Hunter & Sigler (1940, p.162) mention a presumption there are no long-term practice effects in span of apprehension.

11) All three tasks show stable individual differences. Perhaps the best example for memory span is the long history of use of memory span tests as subtests on IQ tests for both children and adults. Woodworth (1938, p. 690) reported measured spans of apprehension ranging from 6 to 11; For absolute judgment see Bachelder & Denny (1977a, p. 140) and Pollack & Ficks (1954, Figure 1).

12) All three tasks show pronounced and surprisingly stable developmental effects. See Blankenship (1938, pp. 13-14) for span of immediate memory; Spitz (1967) for span of absolute

[3] Each test (a) assesses individual differences which are not attributable to individual differences in physical capacities and (b) the test items of each test are scorable in terms of the goodness of performance.

judgment; and Fischer (1992), Freeman (1912), and Gelman & Tucker (1975) for span of apprehension (numerosity).

***Miller's Analyses***

Early in his paper (p. 81) Miller expressed his intent to reinterpret the span tasks as measures of information transmission, rather than the traditional interpretations as measures of memory, attention, and psychophysical processes. He went on to introduce two basic concepts of information theory. The *information content* of a stimulus is measured as the $\log_2$ of the number of stimulus alternatives and *channel capacity* is the upper limit on the ability to transmit information.

Miller illustrated his concepts with the absolute judgment data of Pollack (1952, p. 748, Figure 7), the same data plotted in Figure 1:

> [Pollack] asked listeners to identify tones by assigning numerals to them. The tones were different with respect to frequency, and covered the range from 100 to 8000 cps in equal logarithmic steps. A tone was sounded and the listener responded by giving a numeral. 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 the listeners made many mistakes. (Miller, 1956, p. 83)

Figure 2 (after Miller's Figure 1, p. 83) plots transmitted information as a function of input information. Input information is $\log_2$(set size) and transmitted information is measured as $I_t$ (also known as the Shannon-Wiener method; calculation and application in absolute judgment are described by Garner & Hake, 1952)[4]. Set sizes varied from 2 through 14 alternative tones which correspond with information values of 1 through 3.8 bits. The figure shows the channel capacity effect such that transmitted information is equal to input information until an asymptote is reached around 2.5 bits of stimulus information (set size about 5). In Miller's words:

> The amount of transmitted information behaves in much the way we would expect a communication channel to behave; the transmitted information increases linearly up to about 2 bits and then bends off toward an asymptote at about 2.5 bits. This value, 2.5 bits, therefore, is what we are calling the channel capacity of the listener for absolute judgments of pitch. (p. 84)

[4] In my original abstract I explicitly stated that Miller applied the weighted responses method to memory span, but the $I_t$ method to absolute judgment. However, Miller relied on papers by Pollack (1952, 1953) and did not specifically mention $I_t$. Pollack's papers are not clear in indicating whether they used that specific method or not. They mention Garner & Hake (1951) in a footnote, but do not state they used the method. However, Garner (1951, p. 373) seems to indicate that Pollack (1952) did use the $I_t$ method. Regardless, it is readily shown that $I_t$ can be applied to both memory span and span of absolute judgment and the results indicate similarity of limits, not difference.

The $I_t$ method is a complex correlational method which I will characterize briefly to contrast it with the weighted responses method Miller used with memory span. Calculation of $I_t$ begins with a two-dimensional data table as diagrammed in Table I (after Garner & Hake, 1951, p. 447, Table I). The top row are the stimuli. Along the left are the responses. The cells contain $N_{jk}$, the number of events in cell *jk*, that is, the number of times response *j* was given when stimulus *k* was presented. From this chart you calculate cell, column, and row probabilities in a manner reminiscent of analysis of variance. The probabilities are converted to $I_t$ through a series of summations.

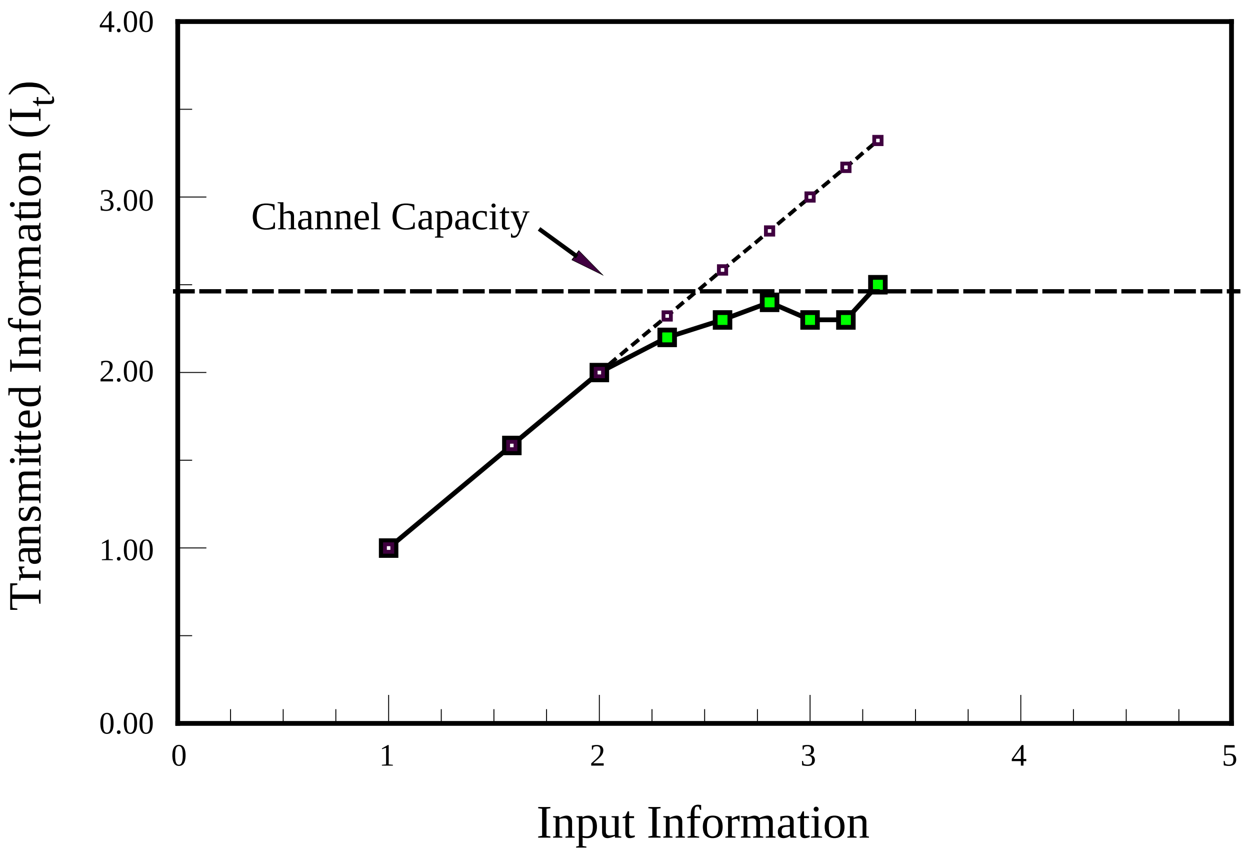


Figure 2: Information transmitted (assessed as $I_t$) as a function of input information in an absolute judgment task. After Miller (1956, p. 83, Fig. 1).

Table I Symbolic Representation of the Number of Joint Occurrences of Stimulus Category (*k*) with Response Category (*j*) (After Garner & Hake, 1951, p. 447, Table I)

| | 1 | 2 | . | $k$ | . | . | $S$ | |
|---|---|---|---|---|---|---|---|---|
| 1 | $N_{11}$ | $N_{12}$ | . | $N_{1k}$ | . | . | $N_{1S}$ | $N_{1.}$ |
| 2 | $N_{21}$ | $N_{22}$ | . | $N_{2k}$ | . | . | $N_{2S}$ | $N_{2.}$ |
| . | . | . | . | . | . | . | . | . |
| . | | | | | | | | |
| $j$ | $N_{j1}$ | $N_{j2}$ | . | $N_{jk}$ | . | . | $N_{jS}$ | $N_{j.}$ |
| . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . |
| $R$ | $N_{R1}$ | $N_{R2}$ | . | $N_{Rk}$ | . | . | $N_{RS}$ | $N_{R.}$ |
| | $N_{.1}$ | $N_{.2}$ | . | $N_{.k}$ | . | . | $N_{.S}$ | $N$ |

$N_{jk}$: The number of events in cell, *jk*.

$N_{j.}$ & $N_{.k}$: The row and column totals.

$N$: The total number of events in the matrix.

***Memory Span.*** After addressing absolute judgment, Miller addressed memory span, arguing:

> We have seen that the invariant feature in the span of absolute judgment is the amount of information that the observer can transmit. . . . If immediate memory is like absolute judgment, then it should follow that the invariant feature in the span of immediate memory is also the amount of information that an observer can retain. If the amount of information in the span of immediate memory is a constant, then the span should be short when the individual items contain a lot of information and the span should be long when the items contain little information (p. 91).

Figure 3 (after Miller's Figure 7, p. 92) plots data in a very different way than does Figure 2. It plots number of items in memory span as a function of information per item, 1.00 bit each for binary digits; 3.32 bits each for decimal digits; 4.70 bits each for letters; 5.04 bits each for letters plus digits 2-9; and 9.96 bits each for 1000 monosyllabic words. As you can see, memory span barely varies as a function of input information measured in this example as $\log_2$(vocabulary size). *Vocabulary* is the same as *stimulus pool* (do not confuse with *stimulus set*).

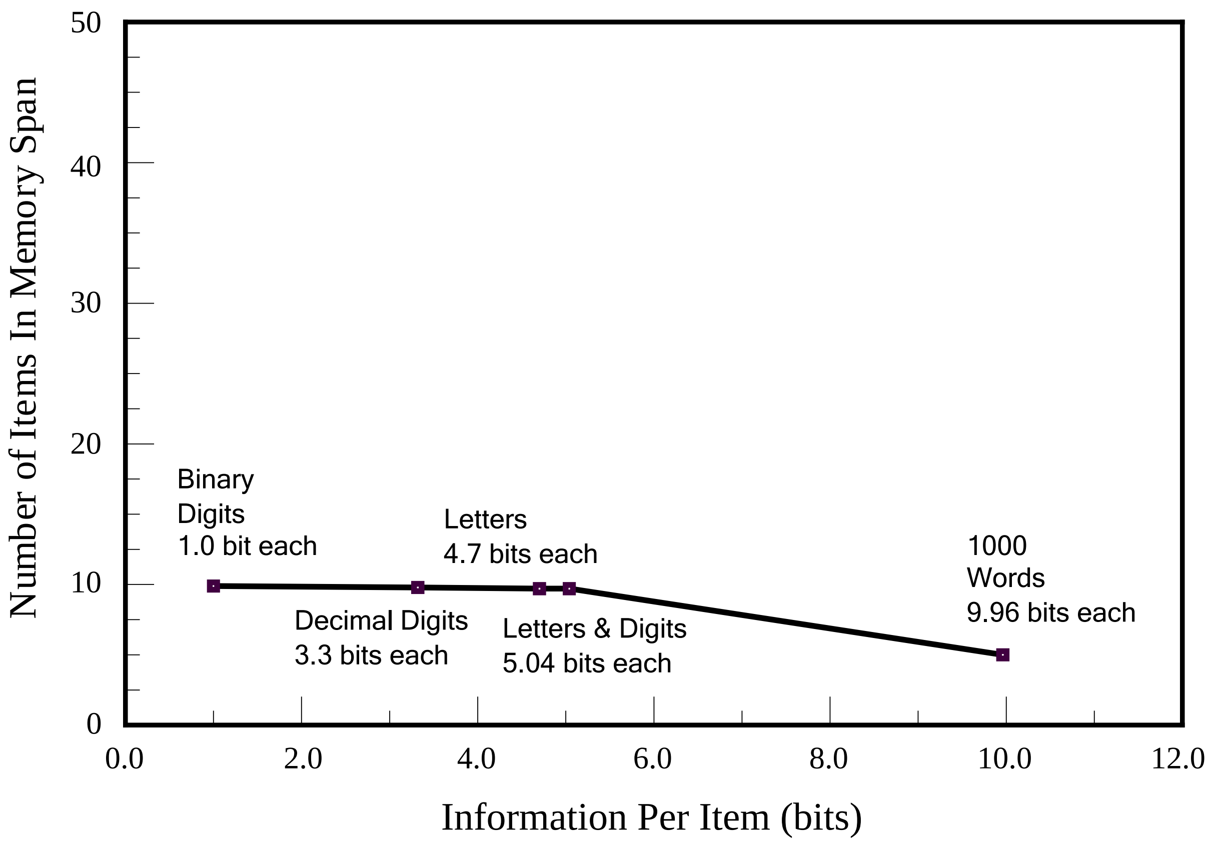


Figure 3. Number of items in memory span as a function of bits of information per item. After Miller (1956, p. 92, Figure 7).

Miller went on to characterize information transmission in memory span assessed by the weighted responses method. He pointed out (p. 91) that we can recall about 7 decimal digits worth 3.3 bits each for a total of 23 bits. Twenty three bits is very much larger than the 2.5 bits channel capacity observed in absolute judgment. While 2.5 bits corresponds to 5.65 alternative stimuli, 23 bits corresponds to 8,388,608 alternative stimuli. Given these results Miller concluded:

> In spite of the coincidence that the magical number seven appears in both places, the span of absolute judgment and the span of immediate memory are quite different kinds of limitations that are imposed on our ability to process information. Absolute judgment is limited by the amount of information. Immediate memory is limited by the number of items. . . .I have fallen into the custom of distinguishing between *bits* of information and *chunks* of information. Then I can say that the number of bits of information is constant for absolute judgment and the number of chunks of information is constant for immediate memory. (pp. 92-93)

***Miller's errors***. To summarize Miller's argument: if the span limits in immediate memory and absolute judgment are the result of the same underlying limitation on channel capacity, then the amount of information transmitted at channel capacity should be the same for each. Since the channel capacity in memory span is so dramatically higher than the channel capacity in absolute judgment, the two limits must have fundamentally different natures.

Miller's argument is flawed in three ways:

1) He stated that the limits in absolute judgment and immediate memory are different kinds of limits because absolute judgment is limited by stimulus information, defined as $\log_2$ (set size), but immediate memory is limited by the number of items. But Figure 1 shows quite clearly that all three spans are limited by set size, and therefore by stimulus information defined as $\log_2$ (set size).

2) He applied two different information measures to absolute judgment and memory span, $I_t$ and the weighted responses method, respectively. It is not at all clear that these two methods are equivalent and Miller offered no evidence or justification for his working presumption. The two methods certainly look very different and the methods of calculation are especially different.

3) He was inconsistent in mapping information concepts on the span phenomena. He assessed stimulus information based on two quite different task characteristics, size of stimulus pool (*vocabulary*) in the case of memory span and size of stimulus set in the case of absolute judgment even though each of the three span tasks can be characterized by both set size and size of stimulus pool and the two task variables are associated with dramatically different experimental effects.

***Channel capacity effect in memory span***. Can we correct Miller's errors by comparing apples with apples? That is, can we apply the $I_t$ method to both memory span and absolute judgment? The answer is yes. It works easier with absolute judgment, but it also works with memory span.

Interestingly, I have a predecessor in this matter. Slak (1976) applied the $I_t$ method to digit span. His result is a curious combination of both the $I_t$ and weighted responses methods and I don't fully understand what he did, but it makes at least part of the point I want to make today: If you apply $I_t$ to both tasks the results indicate the limits are similar, not different.

Figure 4 presents Slak's Figure 1 (p. 494). It plots data for a single subject and clearly illustrates the channel capacity effect in digit span, exactly parallel to the channel capacity effect Miller discussed in absolute judgment (Figure 2). Slak's values for stimulus information and for information transmission are weighted responses measures. Slak (pp. 494-495) pointed out that these values can be converted to "item units" by dividing them by 3.3, the number of bits contained in each decimal digit. Input information in bits is Set Size x 3.3 bits each. The point along the x axis where performance begins to reach asymptote is set size = 6.1, which falls in Miller's 7 +/- 2 range. Similarly, channel capacity in item units in Figure 4 is 6.7.

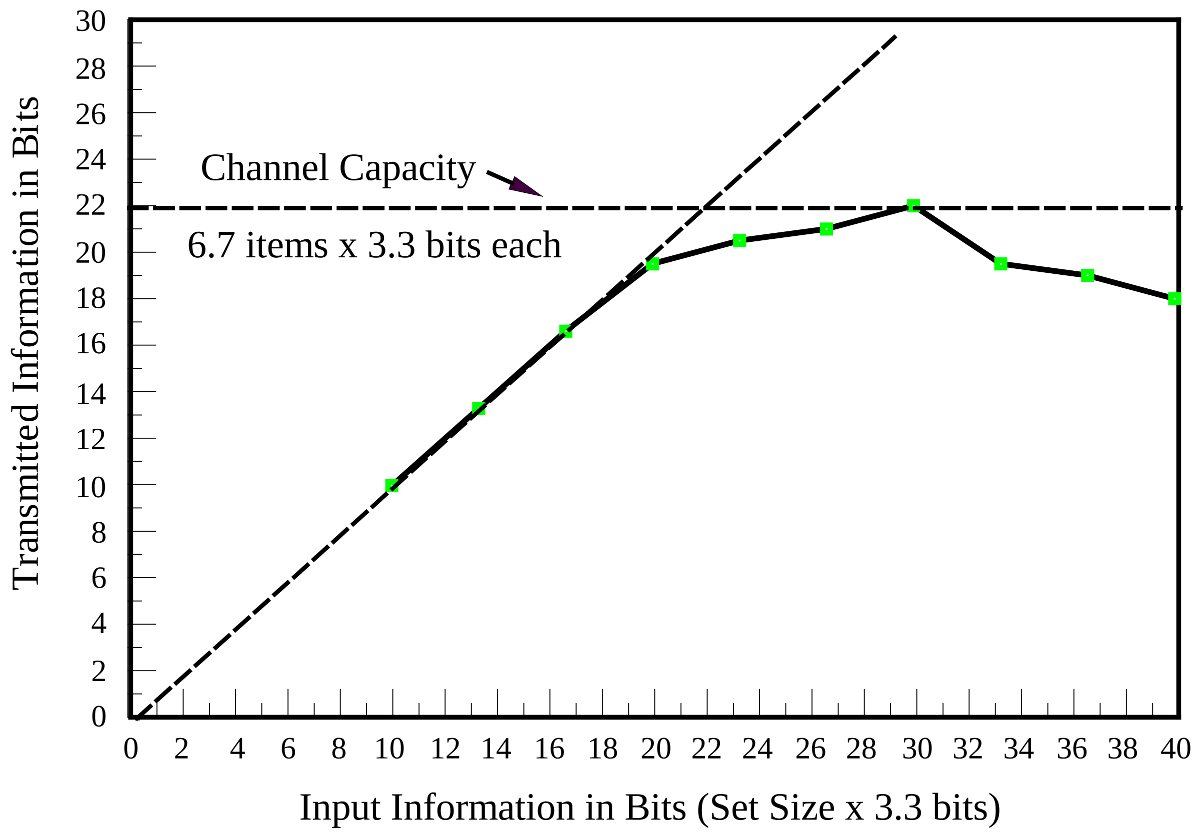


Figure 4. Transmitted information as a function of input information in a digit span task. After Slak (1976, p. 494, Figure 1)

Slak's analysis shows the channel capacity effect for immediate memory, but it does not fully duplicate the channel capacity effect Miller reported for absolute judgment. If $I_t$ could be applied to memory span in a way which more closely resembles Figure 2 for absolute judgment, the present case would be stronger.

I collected some data to illustrate the application of $I_t$ to memory span. My wife tested me with a standard word span test. Since I almost never err on set sizes 1 through 5 we did not test them. She presented 10 stimulus sequences at each set size 6, 7, and 8. The stimulus pool comprised 10 spoken common single-syllable words: *dog, man, tree, cake, fish, chair, school, barn, home, grass*. The rate of presentation was about 1 per sec. The responses were spoken words which she recorded verbatim.

It will help to understand the approach by keeping in mind Miller's definition of channel capacity as "the upper limit on the extent to which the observer can *match* [emphasis added] his responses to the stimuli we give him" (p. 82). The usual measure of memory span counts the number of correct items; the $I_t$ method assesses the match of input with output. Table II presents two of my trials in this word span task, showing the stimulus sequences with the corresponding response sequences. Table II also shows the 2-way data table which is the starting point for calculating $I_t$.

Table II: Stimulus and Response Sequences in a Memory Span Task

A correct response sequence

S: Fish Grass Cake Home Barn School Chair
R: Fish Grass Cake Home Barn School Chair

A wrong response sequence

| | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $S_7$ |
|---|---|---|---|---|---|---|---|
| S: | Fish | Chair | School | Grass | Barn | Home | Cake |
| R: | Fish | Chair | School | Home | Grass | Barn | Cake |

The Data Matrix for Calculation of $I_t$

| | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $S_7$ |
|---|---|---|---|---|---|---|---|
| $R_1$ | // | | | | | | |
| $\mathbf{R}_2$ | | // | | | | | |
| $R_3$ | | | // | | | | |
| $R_4$ | | | | / | / | | |
| $R_5$ | | | | | / | / | |
| $R_6$ | | | | / | | / | |
| $R_7$ | | | | | | | / |

The top pair of S and R sequences is a correct trial; the responses all match the stimuli and the seven S-R events are tallied along the diagonal. The second pair of S and R sequences is a wrong trial. Matches are correct for the first 3 stimuli so they are tallied along the diagonal as shown. For the 4th stimulus I gave the 6th response so the tally is entered in $S_4R_6$. For the 5th stimulus I gave the 4th response so the response was tallied in $S_5R_4$ and so on. To calculate $I_t$ you proceed just as you do for absolute judgment. The results are shown in Figure 5 which shows the same channel capacity effect seen in absolute judgment and the channel capacity is about the same magnitude, 2.5 bits.

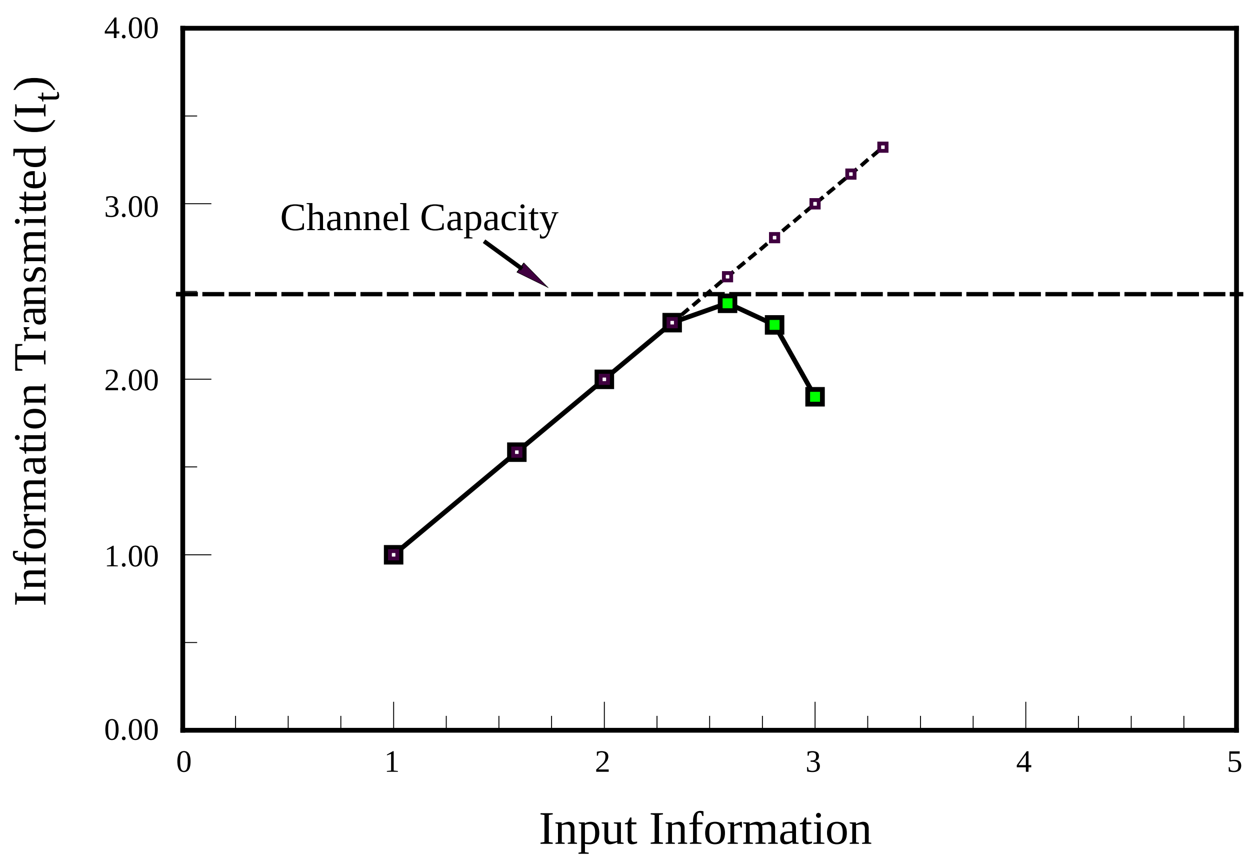


Figure 5: Information transmitted ($I_t$) as a function of input information in a word span task.

For my present purpose, to simply show that $I_t$ can be meaningfully applied to memory span, I scored only response sequences with transposition errors, the most common error in serial recall. Other errors are common, however, including omissions and intrusions. Sometimes the response sequence has fewer items than the stimulus sequence, sometimes more, making scoring challenging. I believe this simplified demonstration makes the point I want to make. I am also confident that experimental procedures and scoring methods can be devised to make the point without discarding excessive numbers of response sequences.

## Summary and Discussion

The spans question (Bachelder, 2005) is one of the oldest and most important questions in experimental psychology. It bears on fundamental working presumptions of both cognitivism and behaviorism and it bears on our understanding of the nature of intelligence. The multiple spans interpretation is the traditional working presumption and the major cognitive models of the three span tasks are based on this presumption.

Miller's paper seems to bolster the multiple spans interpretation (Baddeley, 1994; Shiffrin & Nosofsky, 1994), but his mathematical arguments are flawed. When the errors are corrected, the limit in memory span looks no different than the limit in absolute judgment. Both curves show the channel capacity effect with channel capacity of about 2.5 bits.

This presentation does not prove the span limits are unitary, but by showing the channel capacity effect in memory span it supports the case for a unitary limitation. By identifying and correcting Miller's errors it removes an obstacle to the consideration of unitary hypotheses.

Current cognitive science models do not allow for an underlying unity in the span limits. I believe this is simply because the models start with the presumption of multiple spans and build from there. I am confident that a clever modeler will be able to generate unitary models in the cognitive science style. Cowan's (1988; 1995; 1999; 2001, pp. 91-92) attention model and the span theory concepts of span load and span ability are likely starting points.

If the limits are unitary it brings current cognitive science models into question. Also, since the concept of the chunk was introduced to account for Miller's erroneous conclusion that the two limits are fundamentally different, it brings the concept of the chunk into question.

## References

Bachelder, B. L. (1970/1971). The memory span paradigm: Its use for the analysis of mental retardation. (Doctoral Dissertation, Michigan State University, 1970). *Dissertation Abstracts International, 32*(1), 576B-577B.

Bachelder, B. L. (1977). Memory span tests and the ability to imitate word sequences and sentences. *Ohio Journal of Speech and Hearing, 13*(1), 5-15. (Journal is no longer published. A copy is available from the author.) Submitted for inclusion in *PsycEXTRA*

Bachelder, B. L. (2001). The magical number 4 = 7: Span theory on capacity limitations. *Behavioral and Brain Sciences, 24*(1), 116-117.

Bachelder, B. L. (2005, February). *The spans question: Challenges to the memory interpretation of the "memory" span test.* Paper presented at the Annual Meeting of the North Carolina Cognition Group, Wake Forest University, Winston-Salem, NC. Available from Bachelder or PsycEXTRA

Bachelder, B. L., & Denny, M. R. (1977a). A theory of intelligence: I. Span and the complexity of stimulus control. *Intelligence, 1*(2), 127-150.

Bachelder, B. L., & Denny, M. R. (1977b). A theory of intelligence: II. The role of span in a variety of intellectual tasks. *Intelligence, 1*(3), 237-256.

Baddeley, A. (1994). The magical number seven: Still magic after all these years? *Psychological Review, 101*(2), 353-356.

Binet, A., & Simon, T. (1916). *The development of intelligence in children* (Elizabeth S. Kite, Trans.).Vineland, NJ: Publications of the Training School at Vineland. (Original work published 1905)

Blankenship, A. B. (1938). Memory span: A review of the literature. *Psychological Bulletin, 35,* 1-25.

Cattell, J. M. (1890). V.--Mental tests and measurements. *Mind, 15,* 373-380.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin, 104,* 163-191.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62-101). Cambridge, UK: Cambridge University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87-114.

Dempster, F. N. (1981). Memory Span: Sources of individual and developmental differences. *Psychological Bulletin, 89*(1), 63-100.

Denny, M. R., & Davis, R. H. (1982). *Understanding behavior: Foundations and applications.* Geneva, IL: Paladin.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology* (Henry A. Ruger & Clara E. Bussenius, Trans.). New York: Columbia University, Teachers College. (Original work published 1885)

Fischer, J. -P. (1992). Subitizing: The discontinuity after three. In J. Bideau, C. Meljac, & J. P. Fischer (Eds.), *Pathways to number: Children's developing numerical abilities* (pp. 191-208). Hillsdale, NJ: Erlbaum.

Freeman, F. N. (1912). Grouped objects as a concrete basis for the number idea. *Elementary School Teacher, 12,* 306-314.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution.* New York: Basic Books.

Garner, W. R. (1953). An informational analysis of absolute judgments of loudness. *Journal of Experimental Psychology, 46*(5), 373-380.

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

Garner, W. R., & Hake, H. W. (1951). The amount of information in absolute judgments. *Psychological Review, 58,* 446-459.

Gelman, R., & Tucker, M. F. (1975). Further investigations of the young child's conception of number. *Child Development, 46*(1), 167-175.

Guilford, J. P., & Dallenbach, K. M. (1925). The determination of memory span by the method of constant stimuli. *American Journal of Psychology, 36,* 621-628.

Hamilton, W. (1880). Lecture XIV. In H. L. Mansel & J. Veitch (Eds.), *Lectures on metaphysics by Sir William Hamilton, Bart.* New York: Sheldon. (Original lectures prepared 1836-1837)

Hayes, J. R. M. (1952). Memory span for several vocabularies as a function of vocabulary size. In *Quarterly Progress Report.* Cambridge, MA: Acoustics Laboratory, Massachusetts Institute of Technology.

Hilgard, E. R. (1951). Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 517-567). New York: Wiley.

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review, 75*(3), 242-259.

Hunter, W. S., & Sigler, M. (1940). The span of visual discrimination as a function of time and intensity of stimulation. *Journal of Experimental Psychology, 26,* 160-179.

Jacobs, J. (1887). Experiments on "prehension." *Mind, 12,* 75-79.

James, W. (1950). *The principles of psychology: Vol. 1. The famous long course complete and unabridged* (Authorized ed.). New York: Dover. (Original work published 1890)

Jensen, A. R. (1964). *Individual differences in learning: Interference factor.* Washington, DC: Educational Resources Information Center. (ED 003 482)

Jensen, A. R. (1970). A theory of primary and secondary familial mental retardation. In N. R. Ellis (Ed.), *International Review of Research in Mental Retardation* (Vol. 4, pp. 33-105). New York: Academic Press.

Jensen, A. R. (1986). g: Artifact or reality? *Journal of Vocational Behavior, 29,* 301-331.

Killeen, P. R., & Taylor, T. (2000). How the propagation of error through stochastic counters affects time discrimination and other psychophysical judgments. *Psychological Review, 107*(3), 430-459.

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General, 111*(1), 1-22.

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

Peterson, L. R. (1963). Immediate memory: Data and theory. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes* (pp. 336-353). New York: McGraw-Hill.

Pollack, I. (1952). The information of elementary auditory displays. *Journal of the Acoustical Society of America, 24,* 745-749.

Pollack, I. (1953). The information of elementary auditory displays. II. *Journal of the Acoustical Society of America, 25,* 765-769.

Pollack, I., & Ficks, L. (1954). Information of elementary multi-dimensional auditory displays. *Journal of the Acoustical Society of America, 26*, 155-158.

Shiffrin, R. M., & Nosofsky, R. M. (1994). Seven plus or minus two: A commmentary on capacity limitations. *Psychological Review, 101*(2), 357-361.

Slak, S. (1976). Information-transmission method for expressing short-term memory span as channel capacity. *Perceptual and Motor Skills, 42,* 491-496.

Sperling, G. (1988). The magical number seven: Information processing then and now. In W. Hirst (Ed.), *The making of cognitive science*, (pp. 71-78). Cambridge: Cambridge University Press.

Spitz, H. H. (1967). Information transmission in an absolute judgment task with feedback, using normal and retarded subjects. *Journal of Comparative and Physiological Psychology, 64*(1), 85-92.

Spitz, H. H. (1973). The channel capacity of educable mental retardates. In D. K. Routh (Ed.), *The experimental psychology of mental retardation* (pp. 133-156). Chicago: Aldine.

Staats, A. W., & Staats, C. K. (1964). *Complex human behavior: A systematic extension of learning principles.* New York: Holt, Rinehart and Winston.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence. (4th ed.).* Baltimore: Williams & Wilkins.

Woodworth, R. S. (1938). *Experimental psychology.* New York: Henry Holt and Co.

Wundt, W. M. (1897). Outlines of psychology (Charles Hubbard Judd, Trans.). Leipzig: Wilhelm Engelmann (Original work published 1896) Retrieved February 2, 2005 from http://psychclassics.yorku.ca/Wundt/Outlines/sec15.htm

Wundt, W. M. (1912). *An introduction to psychology.* (R. Pintner, Trans.). London: Allen & Unwin. (Original work published 1911)

## Appendix: The details of Figure 1

The memory span data are from Guilford & Dallenbach (1925, p. 627, Table II). They presented stimulus sequences ranging from 4 to 13 digits, one of each size, to 100 different subjects. The subjects, though not specified, almost certainly were college students. The span of apprehension data are from Mandler & Shebo (1982, p. 8, Figure 3). The only changes made were to convert from probability of an error to probability of a correct response by subtracting Mandler & Shebo's values from 1.0. Finally, the absolute judgment data are from Pollack (1952, p. 748, Figure 7).

Pollack's data had to be corrected to be metrically consistent with the other two curves. This is because, while both the immediate memory task and the span of apprehension task present a full stimulus set on each trial, absolute judgment tasks present a single stimulus on each trial and test a full stimulus set over a block of trials. For example, in a traditional memory span task when a subject hears 7 digits (a full stimulus set) and responds perfectly, the response is scored "correct" for the full stimulus set considered collectively. Any error, even a simple omission of a single digit, is recorded as "wrong" for the full stimulus set. Similarly, in a span of apprehension task a subject sees the full stimulus set, responds with a number, and is scored either "right" or "wrong" with respect to the full stimulus set considered collectively. Span of absolute judgment, however, is usually scored either in terms of the information metric, $I_t$, (which summarizes responding to the full stimulus set considered collectively), or as probability of a correct response to a single stimulus from a full relevant stimulus set.

Pollack presented the same data set both as $I_t$ and as probability of a correct response to a single stimulus (his Figures 3 and 7, respectively). The estimated probability of correctly judging all stimuli of a full stimulus set is simply the product of the individual probabilities, namely, $p' = p^{\text{set size}}$. For example, Pollack's Figure 7 shows a mean probability of about .94 for a correct response to a single stimulus for set size 6. The estimated probability of performing perfectly on the full stimulus set considered collectively is $.94^6$ which equals .69. The corrected values are plotted in Figure 1.