The Memory Span Experiment: A Behavioral Analysis

Bruce L. Bachelder

Independent Practice of Psychology, Retired

Morganton, North Carolina

Author Note

Bruce L. Bachelder, Morganton, NC.

Judith B. Bachelder has done technical reviews and, more importantly, she has made space in our lives, at times under especially difficult circumstances, so I have been able write. For many years the staff of the Burke County Public Library, Morganton, NC have helped me obtain books and articles.

Address correspondence to Bruce L. Bachelder, 306 W Union St, Morganton, NC 28655-3729. 828-430-0421 or 828-205-0041 Fax: NA E-mail: bruce@BruceBachelderPhD.com

Abstract

Performance limitations are observed in three types of experiment known traditionally as experiments on *memory span, attention span,* and *span of absolute judgment.* The three experiments have been used extensively as tools to investigate hypothetical mental and neurological processes and intelligent function, but *span theory* views them in a behavioral way. All three limits are viewed as the same empirical limit on *span ability,* the ability to function under *span load.* Span load is a count of the number of discriminative stimuli that function concurrently to occasion responding; it is determined by task analysis. Span ability is assessed with the staircase span procedure. The theory distinguishes the *response string* from the response chain on the basis of the locus of stimulus control; the response string is controlled by the multiple stimuli in a stimulus string or stimulus complex. Response chains are emitted intraverbals. There is no known limit to the length of a chain that can be attained through extensive experiential history, but the length of the response string is severely limited at around 7 items (the span limit) and is remarkably resistant to change. Span limits increase during the developmental period and covary closely with intelligent function. The values of span ability and span load are potent tools for the prediction and control of developmental and individual differences in diverse experimental paradigms drawn from both behavior analytic and general experimental traditions. The focus at the level of experimental paradigm fosters the unification of cognitive, behavioral, and psychometric traditions.

Keywords: memory span, span of absolute judgment, attention span, intelligent function, intellectual disability, unification

The Memory Span Experiment: A Behavioral Analysis

The span limitations are observed in three types of experiment, known traditionally as experiments on *memory span, attention span*[1], and *span of absolute judgment.* All three span experiments have been used extensively for decades as tools to investigate hypothetical mental and neurological processes and intelligent function. However, the present article focuses on the memory span limit and aims to lay the foundation for presentations of the other two experiments in later articles (see Bachelder & Denny, 1977a, 1977b for an earlier approach to all three). My approach to span limits is referred to as *span theory.* At the core of the theory is a simple elaboration of the basic notion of stimulus control, namely, that description, prediction, and control necessarily involve specification of multiple stimuli functioning concurrently to occasion responding. Furthermore, there are reliable and stable developmental and individual differences in the ability to cope with multiple stimuli. Taking these developmental and individual differences into consideration greatly increases our ability to address, predict, control, and understand developmental and individual differences in levels of functioning in diverse types of complex behavior in experiments drawn from both the behavioral and the general experimental literatures.

In summary, my purposes are (a) to introduce span limits as empirical phenomena with compelling features; (b) to introduce span theory, my behavioral view of the limits and of the nature of intelligent function and intellectual disability (Bachelder, 1970/1971, 2001; Bachelder & Denny, 1977a, 1977b); (c) to point out implications of the span phenomena for behavior

[1] Attention span is also known as span of numerosity, span of apprehension, and subitization span.

theory and application; and (d) to show how span theory fosters integration and unification of the behavioral, cognitive, and correlational (see Cronbach, 1957) traditions.

Comments on Terminology and Strategy

Investigations of the memory span experiment, from the earliest days of scientific psychology, have been intimately linked with hypothetical notions of mental development and intelligence. The first publications of span theory (Bachelder & Denny, 1977a, 1977b) introduced the theory as a behavioral theory of intelligence. Ironically, citations of those articles have sometimes characterized span theory as a memory theory of intelligence (e.g., Chiappe & MacDonald, 2005, p. 14; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002, p. 164).

This presentation does not use *intelligence* in the sense of an explanatory mental faculty, but rather as a convenient reference to the empirical phenomena underlying the traditional notion of intelligence. Span theory is about span limitations and the implications of a behavioral conception of them. *Span ability* is directly inferred from samples of overt behavior. The empirical features of the span limits are remarkably like the features traditionally attributed to the hypothetical construct of intelligence (Bachelder & Denny, 1977a, pp. 130–134). Therefore, a behavioral understanding of the span limits promises to be a solid foundation for understanding intelligent functioning and for integrating the behavioral and cognitive approaches. One of the goals of span theory research is to explore the extent to which understanding of span ability in interaction with experiential history can produce a comprehensive account of developmental and individual differences in the repertoires sampled by IQ tests.

The strategy of span theory research differs from the strategies of both cognitive and behavior analytic approaches to the span phenomena. Consider the memory span limit. Cognitive theory aims to develop explanations of the limit in overt behavior in terms of a limit in hypothetical underlying mental functions. For example, the memory span limit of 7 is sometimes explained as the result of a short-term memory storage capacity limited to seven items. Similarly, behavioral accounts of the limits have tended to explain them in terms of experiential history conceived to control the length of acquired response chains (e.g., Staats, 1968, pp. 405–407).

Span theory pursues a different strategy. It treats the memory span limit, not as a phenomenon to be explained, but as a measurable intelligence-like developmental and individual differences variable useful in prediction and control of behavioral events in diverse experimental paradigms. The intelligence-like features are *empirical* features, not hypothetical features. Span theory research focuses on evaluating the equation, $\text{Performance}_{\text{paradigm}}$ = f(span load, span ability) where (a) $\text{Performance}_{\text{paradigm}}$ refers to performance in a specific type of experimental paradigm, (b) span load is a count of stimuli functioning concurrently to occasion responding, and (c) span ability is a measure of the ability to cope with span load.

The Empirical Basis for the Notion of a Span Limit

The span limit phenomenon is a feature of a simple memory span procedure commonly involving, but not restricted to, the repetition of series of spoken digits. Textual digits, spoken or textual words, pictures, and color samples are also used. The procedure may be viewed as (a) an experiment evaluating the effect of *set size,* the number of digits the subject attempts to repeat; or (b) as an *assessment* in which an examiner aims to assess the span limit viewed as a developmental and individual differences variable. Over the years a number of variations on the same basic procedure have evolved. Here, I'll describe three: (a) the ascending series assessment method, (b) the performance x set size method, and (c) the staircase span ability assessment method.

The Ascending Series Assessment Method

This is the simplest method and a variation of it is used for the *Digit Span* subtests of the Wechsler intelligence scales. The examiner instructs the subject to listen to series of spoken digits then to say them back perfectly; that is, in the same serial order with no omissions or intrusions. The number of digits (set size) presented on a trial varies trial by trial according to the subject's performance. The first trial presents a small set size, for example, 3 8 5. If repetition is perfect, the next trial presents a *different* random sequence one digit longer, for example, 6 9 3 2. Set size increases trial by trial until the subject starts to err (omit, transpose, or intrude items). On error, the next trial presents a different random series of the same set size just failed. If this series is repeated perfectly, the next trial presents a different random series one digit longer. Testing continues this way until the subject fails twice consecutively at the same set size. The largest set size repeated perfectly at least once is taken to be the *digit span* or memory span. Table 1 presents typical spans derived from this procedure.

The memory span procedure must not be confused with a serial learning procedure in which the stimulus sequence is the same trial by trial. Under those conditions subjects easily come to repeat *supraspan* stimulus sequences. Construction of memory span assessment materials avoids sequences such as historical dates; counting sequences such as, *12345;* and alphabetical sequences, such as, *abc*; because they are acquired sequences that through interactional history have become easy to repeat. Such sequences have long been considered to inflate estimates of the "true" memory span limit. From a behavioral point of view these sequences are chains or intraverbals and can greatly exceed the memory span limit.

The Performance x Set Size Method

This procedure is not a common one, but it is fundamental to the span theory point of view. Stimulus test series are constructed just as described above, but performance is assessed for each of a wide range of set sizes, from well below to well above expected span limits. Set sizes 4 through 10 would be appropriate for college students; 1 through 15 would accommodate nearly all human subjects regardless of age and functioning level. For each set size a fixed number of different random series are prepared and each subject attempts all series of each set size. Each trial is scored dichotomously, that is, perfect repetition or not. Overall performance is assessed as the relative frequency of perfect repetition plotted against set size as shown in Figure 1.

Figure 1shows that set size is a potent variable: Performance is an inverse sigmoidal function of set size. Consider the curve for college students. Over small set sizes 1 through 4 or 5 stimulus control is near perfect then deteriorates rapidly over larger set sizes.

The shape of this curve closely resembles the curves encountered measuring psychometric thresholds. This fact gave rise (e.g., Guilford & Dallenbach, 1925, p. 626) to measurement of a span limit as the .50 threshold along the set size dimension. Span theory views the memory span not as the largest correctly produced sequence or the number of items which can be stored in short-term memory; but, rather, a threshold, a mathematical abstraction that estimates the set size that corresponds to a .50 relative frequency of productions for which stimulus control is perfect. The threshold measure of the span limit is different from the largest set size measure, but they are similar in magnitude and closely correlated. The largest-set-size measure can be viewed as a shorthand description of performance in the performance x set size method.

Table 1

Typical Memory Spans Assessed With the Ascending Series Method With Adults and Children With Average or Below-Average IQ

Subjects (IQ)	Span (Range)	Ages	*n*	Source
Old adults[a] (Avg)	6.25 (NA)	75-96	40	Ryan, Lopez, and Paolo (1996, Table 2)
Adults[b] (Avg)	6.71 (2-9[c])	16-69	1800	Gignac and Weiss (2015, Table1 1)
Preschoolers (Avg)	3.0 (2-4)	3.4-4.9	12	Bachelder Files
Adults (81.32)	4.0[d]	16-69[d]	50	Gignac and Weiss (2015, Table1 1)
Adults (57.58)	3.0[d]	16-69[d]	12	Gignac and Weiss (2015, Table1 1)

Note. Avg = Presumed Average IQ. NA = Not available. Except for the preschoolers, stimuli were spoken digits presented at a rate of 1 per s; responses were spoken digits. For preschoolers the stimuli were spoken common words (e.g., farm, boy, doll) presented at a rate of 2 per s; responses were spoken words.

[a]These are individuals with 12 or more years of education selected from the old-age normative sample for *Wechsler Adult Intelligence Scale-R*. [b]This is the normative group for *Wechsler Adult Intelligence Scale-IV.* [c]9 is the largest set size tested; the true range would be larger. [d]These are subjects selected for digit spans of 4 or 3, subsets of the full normative group of 1800 subjects ages 16 through 69.

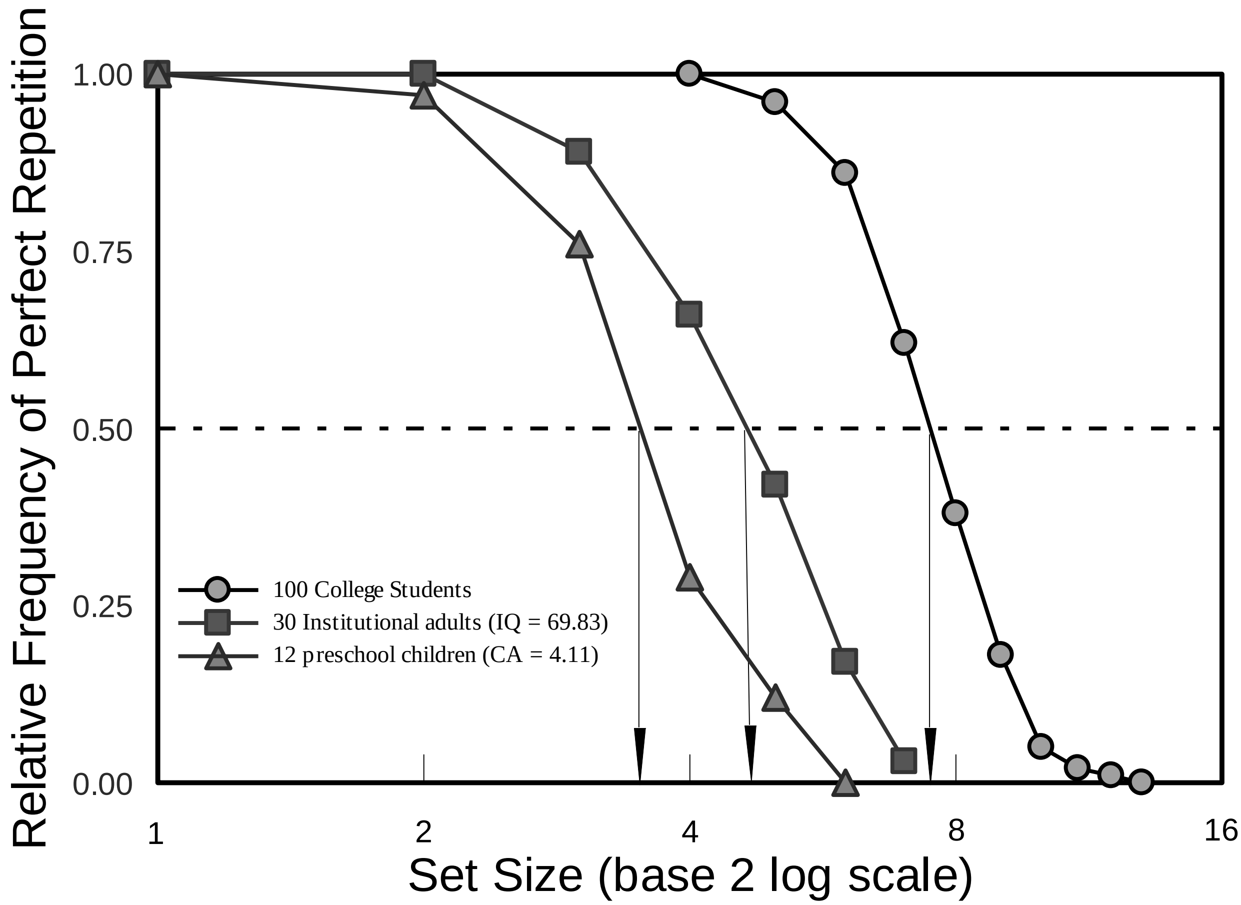

Fig. 1 Relative frequency of perfect repetitions as a function of set size (number of discriminative stimuli on a trial) for college students, institutional adults with low IQs, and preschoolers with normal development. These are group data, but individual data resemble them closely. The three point out the set sizes corresponding to .50 relative frequency. The data for the right-hand curve were taken from Guilford and Dallenbach (1925, Table 2). The data for the middle curve were taken from raw data sheets collected pursuant to Bachelder (1970/1971, Experiment 1). The left- drop lines hand curve presents previously unpublished data from my files.

The Staircase Span Ability Assessment Method

The performance x set size method uses many trials most of which tell us little about the individual subject's span threshold. For example, for college students success at set sizes 1 through 3 and failure at set sizes 8 through 10 do not discriminate span thresholds of 5 through 7. The staircase span assessment method[2] (e.g., Bachelder, 1970/1971, pp. 58–60) is efficient because it presents set sizes that vary closely around the individual subject's span threshold and wastes no time and effort with administration of smaller and larger set sizes.

The method has two parts. It begins with the ascending series method and terminates after failure at two successive set sizes. At the end of each trial the examiner states, "That was right." or "That was wrong." as appropriate. The largest number of digits repeated perfectly is taken as a quick estimate of the span threshold. The second part, the staircase trials, presents a fixed number of trials with set sizes varying trial by trial according to the following rules: The set size of the first staircase trial is equal to the largest set size correctly repeated in the ascending series part. Whenever a series is repeated perfectly, the next set size is one digit larger. Whenever a series is failed, the next set size is one digit smaller. Thus, presented set sizes closely bracket the individual subject's span threshold so the number of subspan and supraspan trials is greatly reduced. The mean of the set sizes presented in the staircase trials is a direct estimate of the span threshold. The values are similar to those of the ascending series method, but more reliable. Ten staircase trials produce staircase span thresholds of acceptable reliability (about .80 to .85); more trials can be used to increase reliability to .95 or more (Jensen, 1970, p. 72). Table 2 presents typical memory spans derived from this procedure.

[2] This method is an adaptation of Cornsweet's (1962) method of measuring absolute psychophysical thresholds.

Summary

Subjects listen to series of *random* digits then attempt to say them back in order with no omissions, transpositions, or intrusions. Small numbers of digits are easy. Almost no one, toddler or adult, intellectually superior or disabled, ever fails perfect performance with 1 or 2 digits. Difficulty increases rapidly with increases in set size and few individuals reliably and perfectly repeat 10 or more. The span limit is assessed as the largest set size repeated perfectly or as a threshold, the set size corresponding to a .50 relative frequency of perfect repetition. The size of the span limit is a stable developmental and individual differences variable, remarkably resistant to improvement with practice and training.

Early Undeveloped Behavioral Comments on the Memory Span Experiment

Publications on the memory span experiment are numerous and go back to the earliest days of scientific psychology (e.g., Galton, 1887; Jacobs, 1887). For reviews see Bachelder and Denny (1977a, 1977b), Blankenship (1938), Cowan (2001), and Dempster (1981). There has been general agreement on two points: (a) the basic procedure that defines the limit; and (b) the empirical fact that measured span limits bear multiple, surprising, and strong relations with developmental and individual differences in the behavioral phenomena referred to traditionally as wit, intelligence, mental ability, mental development, competence, functioning level, complex behavior, cognitive behavioral repertoires, intellectual disability, and the like. The memory span literature is huge and largely mentalistic; but there have been occasional comments from behavioral perspectives.

Skinner (1957, p. 64) discussed “digit repetition” in *Verbal Behavior*. He characterized it as echoic verbal behavior that he discussed as critical for the development and teaching of basic verbal repertoires (pp. 56, 62). He spoke of echoing behavior as an acquired skill (1957, pp. 56–

Table 2

Typical Staircase Word Spans for Several Populations Varying Widely in Functioning Level

Subjects	Span (Range)	Ages	*n*
College Students	6.01 (4.1-7.9)	--	20
Staff of a residential facility	5.40 (4.3-6.9)	--	20
Kindergarteners	4.30 (3.3-5.0)	5.83	22
Preschoolers	3.63 (2.5-4.7)	4.11	12
Facility residents (IQ = 46.0)	3.06 (1.5-4.7)	21.50	18

Note. These are previously unpublished data collected under conditions I take as standard for measuring span ability: (a) The materials were spoken single-syllable words (e.g., farm, boy, doll) judged to be well established in subject repertoires; (b) the rate of presentation was 2 words per s; and (c) responses were spoken. Except for the college students and the facility residents, subjects were presumed to be of average functioning levels. The facility was a state residential facility for intellectually disabled children and young adults. The kindergarteners were enrolled in a church kindergarten and the preschoolers were enrolled in a community day care.

58) and mentioned that “the length of the verbal stimulus which can be successfully echoed is sharply reduced in some cases of aphasia” (p. 64).

Bachelder (1970/1971) addressed the multiple stimuli in memory span experiments via the construct of *elicitation span,* conceived to be a limit on the number of stimuli that can evoke a series of responses. He introduced the construct of *response string* to discriminate it from the response chain. The response string is controlled directly by the memory span stimulus sequence, that is, the stimulus string. A response chain results from practice and does not depend on an antecedent stimulus string. Bachelder speculated that (a) developmental and individual differences in elicitation span interact with experiential history to give rise to developmental and individual differences in acquired repertoires, and (b) the development of teaching methods and goals that take span limits into account promise to improve the efficiency and efficacy of teaching and training.

Mahadevan, Malone, and Bailey (2002, pp. 9–10) interpreted the digit span experiment in terms of tacts and intraverbals. They focused, not on the span limit, but on the subject’s development of an extensive set of tacts highly effective in repeating digits. However, at one point they alluded to the span limit as conceived in span theory: “[The subject]found that he could produce a string of two intraverbals (i.e., two chains of tacts) but any attempt to string more than four tacts within an intraverbal led to errors of transposition in the response” (p. 9). This is a good description of the span limit as a behavioral phenomenon illustrated in Figure 1: Performance is essentially perfect up through set size 4 with errors starting with set size 5. Even though the subject developed a repertoire of tacts effective in repeating large numbers of digits, he could not reliably emit more than four such tacts in one verbal episode.

Compelling Features of the Memory Span Limit

As a measure of the hypothetical construct of intelligence the memory span procedure has poor *face validity,* but it has surprisingly good *construct validity.* That is, the memory span procedure doesn't look like it could measure intelligence, but memory span thresholds have the empirical features traditionally attributed to intelligence:

- Span thresholds increase during the developmental period and reach asymptote in early adulthood, closely paralleling the course of mental age changes and physical development.
- At each age there are stable individual differences in the span thresholds.
- Span thresholds are smaller than average among children and adults with intellectual disability and larger than average among children and adults with superior intellectual functioning.
- The span thresholds are highly reliable when properly measured.
- Practice and training are remarkably ineffective in increasing span thresholds.
- Students at the bottom of their class by teacher judgment have smaller span thresholds.
- The relationship of memory span thresholds to IQ scores is moderate to high.
- The complexities of language imitation and language utterance covary closely with span thresholds.
- There is no gender effect.
- There is no race effect.
- There is little or no culture effect.
- There is no Flynn effect (Unlike IQ, span thresholds have not increased over the years)

It is especially significant to point out that these features of intelligent function are deliberately built into IQ tests, but the memory span experiment was not developed that way. These features are “natural,” that is, they are inherent in the memory span experiment.

Span Theory

Span theory arose (Bachelder, 1970/1971) in a behavioral tradition after a serendipitous encounter with the memory span subtests of the Wechsler and Stanford-Binet IQ tests. The span phenomena impressed me as anomalous from a behavioral point of view, suggesting that an experimental analysis would be fruitful for development of basic theory. Additionally, because there is such a potent relationship between size of the memory span threshold on the one hand and developmental level and status as normally developing or intellectually disabled on the other; investigation of the span limit experiments promised to be fruitful in efforts to understand intellectual disability so as to develop improved methods of teaching and training.

The Domain and Constructs of Span Theory

Domain. Span theory focuses on humans functioning in diverse experimental paradigms. The primary, but not exclusive, focus has been on the three span limit experiments. The focus of this section is the staircase memory span assessment procedure. The procedure operationalizes the construct of span ability and serves as the method for assessment of developmental and individual differences in it.

The analysis presumes subjects come to the session having had extensive histories in home and classroom settings where they engaged in close instruction-like interaction with parent or teacher figures. This complex interactional history constitutes *setting conditions* such that (a) subjects listen and follow instructions carefully; and (b) when the examiner speaks any one of the 10 digits the subject repeats the digit immediately and with a likelihood approaching 1.0. These presumptions are valid for a wide range of human subjects: children and adults with normal development and children and adults with intellectual disabilities.

The staircase span procedure comprises three parts: (a) instructions, (b) the ascending series, and (c) the staircase trials. The instructions constitute setting conditions such that the subject listens for and repeats digits. The ascending and staircase trials comprise multiple verbal episodes wherein the examiner utters a spoken digit and the subject responds with the corresponding spoken digit. The verbal episodes may be viewed as mands, that is, the examiner emits digit words that specify responses to be emitted by the subject. A more accurate view is in terms of rule governed behavior. The examiner emits verbal responses in accordance with rules of the staircase span method and the subject emits responses in accordance with the rule: Say the same digits the experimenter does. At the end of the procedure the examiner calculates the mean set size presented over the 10 staircase trials. Thus, the final span threshold score is a mathematical abstraction that estimates the .50 span thresholds illustrated in Figure 1.

Constructs. This section discusses the *stimulus string*, the *response string* (distinguished from the response chain), *span load*, and *span ability.* Discussion relies on Table 3 which symbolizes what happens on single trials in memory span experiments. The presentation makes use of Skinner's notion of the *verbal episode* (Skinner, 1957, p. 2) comprising one verbal exchange between a speaker and a listener. Each trial of a memory span experiment comprises two verbal episodes. In the first verbal episode the experimenter says digits and the subject responds with spoken digits. In the second verbal episode the experimenter consequates the response with an intraverbal (S^{rein}), "That was right" or "That was wrong." The span theory analysis focuses on the first verbal episode of each trial.

Consider a trial with set size 1: The experimenter says "2" and the subject responds with "2." This can be interpreted and symbolized in the usual way in terms of the three-term contingency: The experimenter’s response, “2,” is a verbal discriminative stimulus (S^{DV}) that occasions the verbal response, “2” (R^{V}).

To this point in my analysis there is nothing new. It is a fundamental working presumption of behavior analysis that one stimulus can function to occasion one response. Consider a trial with set size 3 as shown in the second example in Table 3. The trial starts with a *single* verbal episode comprising (a) three discriminative verbal stimuli, “4 9 2;” that (b) function concurrently to occasion three verbal responses, “4 9 2.” The span theory interpretation at this point departs from the common behavioral interpretation of the series of responses as a chain of associated responses. Instead, span theory asserts: S_1^{V} (“4”) occasions R_1^{V} (“4”), S_2^{V} (“9”) occasions R_2^{V} (“9”), and S_3^{V} (“2”) occasions R_3^{V} (“2”); that is, verbal discriminative stimuli 1 through 3 function *concurrently* to occasion verbal responses 1 through 3. Given these interpretations, the span theory term for the span limit is not memory span; but, rather, the *span of discriminative stimuli*.

The third example in Table 3 is the generalized expression for set size n. Figure 1 plots performance as a function of set size (n). For college students stimulus control for set sizes 1 through 4 is essentially perfect. However, at set size 5 performance starts a precipitous decline through .50 to .00 relative frequency. The set size corresponding to .50 relative frequency correct is taken as the measure of span ability.

Table 3

Symbolization of trials for set sizes 1, 3, and n (the general case) in a memory span experiment

Set Size (span load)	Stimulus String	Response String	Symbolization		
1	2	2	S^V	R^V	S^{rein}
3	4 9 2	4 9 2	$S_1^V\ S_2^V\ S_3^V$	$R_1^V\ R_2^V\ R_3^V$	S^{rein}
n	8 5 . . . 1	8 5 . . .1	$S_1^V\ S_2^V \ldots S_n^V$	$R_1^V\ R_2^V \ldots R_n^V$	S^{rein}

Note. Set size (span load) = the number of stimulus items on a specific trial. Stimulus String = examples of digit sequences. Response string = examples of response sequences. S^V = verbal discriminative stimulus, R^V = verbal response, S^{rein} = reinforcing stimulus. All three examples, for set sizes 1, 3, and n; are considered single verbal episodes varying in set size (complexity).

The series of S^Vs on a trial is called a stimulus string or stimulus complex. The series of R^Vs is called a response string (Bachelder, 1970/1971, p. 2; Bachelder & Denny, 1977a, pp. 137–138). The distinction of response string from response chain is based on the nature and locus of stimulus control: Stimulus control for a response string resides in the stimulus string: The individual stimuli function concurrently to occasion the individual responses in the response string. In contrast, the traditional notion of a chain has it that stimulus control resides in the chain itself, that is, the first response occasions the second, the second occasions the third, and so on.

The notion of a span limit applies to the stimulus string and response string considered as a single verbal episode. As shown in Figure 1 the length of a response string is severely limited. In general, response chains are developed out of response strings. Consider a young boy learning the alphabet. When first instructed to recite he can't, simply because he has not had the requisite experiential history. He is, however, able to repeat the stimulus string, "a b c," uttered by his parent or teacher. His early responses, "a b c," are response strings controlled by the stimulus strings. After much practice the student develops a response chain for which the prompting stimulus strings of the early trials are no longer necessary. Response strings are limited in length, but are evoked immediately. Response chains have no known limit, but require practice for their development and long chains require extensive practice.

Set size is a term common in the cognitive science literature. It is useful both there and here as *descriptive* of procedure. The span theory term corresponding to set size is *span load*.[3] Span load is an objective feature of an experimental paradigm and is determined through task analysis. It is a count of the number of S^Vs that function concurrently to occasion the response string.

[3] The corresponding cognitive term for set size is *memory set*, viewed as the set of mental representations held in short-term memory as a basis for production of responses.

It might be argued that the notions of the response string and span ability are unnecessary because interpretation of response sequences as chains is parsimonious and squares with everyday and laboratory experience. A full discussion is beyond the scope here, but I can sketch my analysis:

- Estes (1974, p. 742) stated, “A substantial body of research and theory … shows quite clearly that this [chaining] interpretation … is inadequate and is in fact quite possibly entirely wrong.”
- The span limit is observed, not just in the memory span experiment involving response sequences, but also in the attention span and span of absolute judgment experiments that do not. An explanation of the span limits in terms of a general and unitary ability to cope with multiple stimuli applies to all three span limit paradigms so it is not only parsimonious, it integrates three research domains traditionally considered to be distinct.
- The values of the span limits are so remarkably similar in diverse experimental paradigms and for diverse stimulus and response repertoires that it strains credulity to presume that developmental and individual differences in the limits are mere reflections of developmental and individual differences in experiential history and the development of chains.

It might be argued that the span limit is the direct result of a decrement in response tendency as a function of time between stimulus presentation and response. Two empirical facts challenge that interpretation. First, errors occur at their highest rate for the stimuli in the middle of the stimulus string, not for the earliest presented items. Second, the same span limit is observed if stimuli are presented simultaneously (visual stimuli).

Summary and Implications for Theory, Practice, and Unification

Summary

I have introduced (a) the span limits as behavioral phenomena with compelling features and (b) span theory, a behavioral view of the limits:

The span limit. The span limit is a limit on function under variation of set size. The span limit is a limit on the response string, not the response chain. The empirical features of the span limits are remarkably similar to the features traditionally attributed to the hypothetical notion of intelligence. This notion of intelligence as a mentalistic explanatory construct has no place in behavioral analyses, but the *features* underlying the notion are important phenomena meriting close attention.

Span theory. Span theory focuses its analyses on humans functioning in diverse experimental paradigms drawn from both general experimental and behavior analytic traditions. Span load is a critical feature of an experimental paradigm; it is determined by task analysis. Span ability is the ability to function under span load; it is measured by the staircase span ability assessment method. Humans differ markedly in span ability, both developmentally and individually, and those differences are remarkably stable. When brought to bear by a skilled experimenter or theorist, the values of span load and span ability function potently in prediction and control.

Implications for Basic Theory

The three-term contingency. This fundamental relationship is usually characterized as involving a single discriminative stimulus, a single operant response, and a reinforcing stimulus. One implication of span theory is that in humans this unit of analysis must be conceived to vary in span load (complexity), that is, in the number of discriminative stimuli functioning concurrently to occasion responding. Furthermore, there are stable developmental and individual differences in the ability to cope with span load.

Stimulus string and response string. Span theory distinguishes a response string (Table 3) from a response chain. A response string and a chain may appear formally identical, but they are functionally different because a response string is occasioned by a stimulus string, that is, multiple S^Ds functioning concurrently. A response chain is a sequence acquired through a history of practice and reinforcement and doesn't require a complex stimulus string for its emission. Consideration of the relation between response strings and response chains promises to clarify the relation between developmental and individual differences in span ability and the acquisition of complex repertoire.

Practice and training. Practice and training have potent effects on the development of response chains, but surprisingly weak effects on the response string. A detailed discussion is beyond the scope here, but the issue is critical for behavioral metatheory and I'll sketch my view of it. Research on practice and training on the memory span limit, by and large, has not clearly discriminated the difference between response strings and response chains. It is obvious that practice and training have potent effects on the development of response chains for almost all higher organisms. The response string, however, is a different matter. At the very least it is surprisingly unresponsive to practice and training. An experiment on "expert memory" by Ericsson and Chase (1982) illustrated this well. One undergraduate student repeated digits "1 hour a day, 3 to 5 days a week, for 20 months" (p. 608). "During the first 4 hour-long sessions his "memory span stayed within a normal range of 7 to 9 digits" (p. 609). This observation illustrates the remarkable stability of the memory span limit in the face of practice.

After the first 4 sessions the subject's behavior started to change and he began to develop a personal and highly effective method of engaging digits based on his extensive experiential history as a runner and his extensive ongoing history repeating digits. Eventually, he was able to repeat “around 80 digits” (p. 609). However, this improvement in digit repetition did not transfer to a different material. His memory span for consonants “remained at six” (p. 610). The span theory interpretation is that the experiment showed that the subject developed a complex repertoire of response chains that enabled repetition of long digit sequences (this interpretation is similar to the interpretation suggested by Mahadevan et al., 2002, pp. 9–10). The fact that the span for consonants did not change is interpreted in span theory as a failure of the experiment to change span ability. Mahadevan et al. (2002) did not comment on this finding.

Ability constructs. Behavior analysis does not easily embrace ability constructs. Skinner (1947/1972, pp. 235–236) seemed to taint the very notion of an ability. He discussed it in context with mentalistic notions such as wants, faculties, and capacities: “Some of them, like wants and attitudes, come to us trailing clouds of psychic glory … others, like abilities and traits, have been made respectable through correlational analyses, which give them the status of “individual differences.” Meehl (1986, p. 332) observed, "Operant behaviorists often dislike trait language, but they need not."

The topic bears detailed analysis beyond the scope here, but I can sketch my view. In traditional experimental psychology the span ability construct would be considered to be either a hypothetical construct or an intervening variable (see MacCorquodale & Meehl,1948). I argue that descriptions, constructs, empirical generalizations, and mathematical equations function as rules governing the behavior of the experimenter or theorist in the enterprise of prediction and control. Whether span ability is an inherent feature of behavior or an acquired quality is an important question; either way, however, the importance of the span ability construct lies in its utility for prediction and control.

Multiple control. The memory span experiment as interpreted in span theory clearly involves multiple stimuli controlling responding. Span theory did not develop in the context of the behavior analytic notion of multiple control, and the precise relation between the two approaches remains to be worked out. Topics we discussed in the first span theory publications (Bachelder & Denny, 1977a, 1977b) overlap considerably with topics addressed as multiple control by Skinner (1957, Chapter 9), Lowenkron (1998), and Michael, Palmer, and Sundburg (2011, p. 13). Overlapping topics include complex stimulus control, intelligent function, conditional discrimination, behavioral complexity, language, reading, cognitive processes, the significance of imitative or echoic prompts in the acquisition of new repertoire, and the emergence of novel behavior. None of these behavior analytic articles mentioned any of the three span limit experiments, all of which involve multiple controlling stimuli.

To the best of my knowledge, behavior analysis has no explicit term for span load (concurrent stimulus control) as I have used it; but there have been allusions to the phenomenon. Schlinger and Blakely (1994, p. 46) stated, "in a conditional discrimination, the evocative functions of one stimulus depend on the presence of other stimuli" (p. 46). Michael et al. (2011, p. 13) made the same point. The notions of joint control (Lowenkron, 1998) and multiple control (Skinner 1957, Chapter 9) appear to be highly relevant, but none of these articles mentioned a span limit effect or any of the three span limit experiments.

Implications for Practice and Application

Intelligent function. Developmental and individual differences in intelligent function have long been explained by the hypothetical notion of intelligence, but the behavioral, size-of-repertoire view has it that they are simply individual differences in total repertoire acquired through diverse interactional histories (Humphreys, 1994, p. 180; Staats, 1996, pp. 271–272). Individuals with intellectual disabilities are considered to be individuals with deficient repertoires that resulted from deficiencies in their interactional histories (Bijou, 1966, p. 2) or from a combination of the effects of biological abnormalities and environmental causes (Staats, 1996, pp. 271–272).

Span theory elaborates on the size-of-repertoire notion of intelligent function. Developmental and individual differences in span ability bear directly on function in diverse settings and diverse experimental paradigms, thus they bear directly on our understanding of the acquisition of new repertoire. Consider experimental paradigms with small span loads to be *simple* and experimental paradigms with large span loads to be *complex.* Individuals with smaller span abilities function well, even normally, in simple experimental paradigms such as the free operant, the discriminated operant, the simple two-choice discrimination, and the conditional discrimination paradigms. Individuals with larger span abilities function well in both simple and complex paradigms, so they can acquire larger total repertoires with larger proportions of complex functions.

Teaching and training individuals with intellectual disabilities. One of the major successes of behavioral psychology resulted from the application of basic principles and procedures to enhance the behavior repertoires of children and adults with intellectual disabilities. Working with skilled behavioral specialists, individuals with intellectual disabilities have acquired repertoires once considered impossible. Some of this success has had to do with sophisticated contingency management; span theory has little to say about that. Much of the success of a behavioral approach, however, has had to do with sophisticated management of stimulus control. Task analysis breaks a complex task, such as putting on a shirt, into a sequence of simple steps which can be acquired separately then chained together. From a span theory point of view breaking a task into simple steps is a matter of analyzing a task with a large span load into a series of tasks with small span loads; that is, the process adjusts task demands to be within the span ability of the individual. In span theory terms acquisition of repertoire is optimized when the match between span ability and span load is optimized.

Discrimination learning. One of the early successes of span theory (Bachelder & Denny, 1977b, pp. 238–240) was an account of the relative difficulty of simple, conditional, and oddity discrimination learning. Almost any subject, infrahumans and humans of almost any age or functioning level, can acquire a simple discrimination. Conditional discriminations are more difficult and lower-functioning humans can have considerable difficulty with them. Oddity discriminations are even more difficult. Below a certain mental age subjects find oddity discriminations essentially impossible. Task analysis of the three discrimination experiments yields span loads of 1, 2, and 3 for the simple, conditional, and oddity experiments, respectively. These three different span loads account for the relative difficulty of the three experiments as well as the interaction of functioning level and type of experiment. Smaller span individuals can learn simple discriminations but struggle or fail on conditional and oddity discriminations. Larger span individuals cope effectively with all three types of discriminations.

Implications for Integration and Unification

Integration. Integration of subdisciplines is always a desirable feature of a theory and a sign of a maturing science. It has long been presumed that the three span limits are fundamentally different. Accordingly, the three span experiments have been used to investigate hypothetical processes in the separate subdisciplines of memory, attention, and psychophysics. Span theory views all three limits as variations on a unitary limitation on span ability. This unitary conception of the three limits links the traditional subdisciplines of experimental psychology and promises to be the basis for their integration.[4]

Unification. The three span limits viewed in terms of span theory lie at the intersection of cognitive, behavioral, developmental, and psychometric traditions. Understanding the limits promises to bridge and unify these traditions:

- It has become conventional wisdom (e.g., Miller, 2003, p. 142) that behaviorism constrained the investigation of complex behavior and is inherently incapable of dealing with cognition, intelligence, and mental development; but from a span theory point of view this is not true. Span theory is a theory of stimulus control and "behavior under stimulus control is essentially the field of cognition" (Sidman, 1978/2010, p. 129).
- Span theory focuses at the level of the experimental paradigm, so it greatly expands our ability to bring a behavioral point of view to bear on prediction, control, and understanding of developmental and individual differences in levels of functioning in diverse types of experiments drawn from the general experimental, cognitive, and behavioral traditions. The present article has focused on the memory span experiment, but Bachelder and Denny (1977a, 1977b) also addressed the attention span and absolute judgment experiments; simple, conditional, and oddity discriminations; intelligent function; language; and reading.
- The free operant experiment fits neatly within the span theory point of view: By definition, it involves no discriminative stimulus so the span load is 0. This means that the notion of span ability is almost irrelevant. Almost all subjects emit responses and adapt to the schedule of reinforcement in force. As experiments increase in complexity in

[4] It appears that cognitive science is converging on a unitary conception of all three limits in terms of a limited working memory.

the sense of involving larger span loads consideration of span ability becomes more critical for understanding.

- The performances of nearly all human subjects fall easily on the performance x set size axes shown in Figure 1, so span theory integrates investigation of adults with normal development with investigation in the subdisciplines of developmental psychology and intellectual disabilities.
- Span ability is an ability construct so the methods and concepts of classical and modern test theory have been useful. In this way span theory integrates correlational traditions with the behavior analytic tradition.

Closing Statements

Span theory arose as an attempt to understand the span limitations in behavioral terms. The span theory concepts of span load and span ability appear to be related to the behavior analytic concepts of multiple and joint control. Much work remains to explicate that relationship, but the paper by Michael et al. (2011) surely indicates the two approaches are converging:

> Multiple control is central to human behavior so much so that one can speculate that quantitative differences in sensitivity to many concurrent variables might underlie both species differences and individual differences within our own species. How much of what we call “intelligence” can be more concretely explained as a sensitivity to concurrent variables or as a skill in manipulating them for strategic purposes? How large a role do deficiencies in such sensitivity or skills play in the child suffering from autism or other disabilities? Skinner’s concept of multiple control seems to lie at the heart of the most perplexing questions about human behavior and of our attempts to answer them. (p. 20)

Compliance with Ethical Standards

Funding: Preparation of the manuscript was funded by the author.

Conflict of interest: I have no conflict of interest.

Ethical approval: All span testing procedures were in accordance with the ethical standards of the institutional research committees and with the 1964 Helsinki declaration and its later amendments. Informed consent was obtained for all children and adults and participation was always voluntary.

References

Bachelder, B. L. (1970/1971). The memory span paradigm: Its use for the analysis of mental retardation. Ph. D. Dissertation. *Dissertation Abstracts International, 32*(1), 576B-577B, (UMI No 71–78, 158): Michigan State University. Dept. of Psychology, 1970.

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., & Denny, M. R. (1977a). A theory of intelligence: I. Span and the complexity of stimulus control. *Intelligence, 1*(2), 127–150. doi: 10.1016/0160–2896(77)90001–0

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. doi: 10.1016/0160–2896(77)90008–3

Bijou, S. W. (1966). A functional analysis of retarded development. *International Review of Research in Mental Retardation, 1*, 1–19. doi: 10.1016/S0074–7750(08)60159–7

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

Chiappe, D., & MacDonald, K. (2005). The evolution of domain-general mechanisms in intelligence and learning. *Journal of General Psychology, 132*(1), 5–40. doi: http://dx.doi.org/10.3200/GENP.132.1.5–40

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*(2), 163–184. doi: http://dx.doi.org/10.1016/S0160–2896(01)00096–4

Cornsweet, T. N. (1962). The staircase-method in psychophysics. *American Journal of Psychology, 75,* 485–491. doi: http://dx.doi.org/10.2307/1419876

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24,* 87–185. doi: http://dx.doi.org/10.1017/S0140525X01003922

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *The American Psychologist, 12,* 671–684. doi: http://dx.doi.org/10.1037/h0043943

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin, 89*(1), 63–100. doi: http://dx.doi.org/10.1037/0033–2909.89.1.63

Ericsson, K. A., & Chase, W. G. (1982). Exceptional memory. *American Scientist, 70*(6), 607–615.

Estes, W. K. (1974). Learning theory and intelligence. *American Psychologist*, *29,* 740–749. doi: http://dx.doi.org/10.1037/h0037458

Galton, F. (1887). Supplementary notes on “prehension” in idiots. *Mind*, *12,* 79–82.

Gignac, G. E., & Weiss, L. G. (2015). Digit span is (mostly) related linearly to general intelligence: Every extra bit of span counts. *Psychological Assessment*, pp. 1312–1323. doi: 10.1037/pas0000105

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. doi: http://dx.doi.org/10.2307/1413916

Humphreys, L. G. (1994). Intelligence from the Standpoint of a (Pragmatic) Behaviorist. *Psychological Inquiry, 5*(3), 179–192. doi: http://dx.doi.org/10.1207/s15327965pli0503_1

Jacobs, J. (1887). Experiments on “prehension.” *Mind, 12,* 75–79.

Jensen, A. R. (1970). A theory of primary and secondary familial mental retardation. In N. R. Ellis (Ed.), *International review of research in mental* (pp. 33–105). New York: Academic Press.

Lowenkron, B. (1998). Some logical functions of joint control. *Journal of the Experimental Analysis of Behavior, 69*(3), 327–354. doi: http://dx.doi.org/10.1901/jeab.1998.69–327

MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review, 55,* 95–107. doi: http://dx.doi.org/10.1037/h0056029

Mahadevan, R., Malone, J. C., & Bailey, J. (2002). Radical behaviorism and exceptional memory phenomena. *Behavior and Philosophy, 30,* 1–13.

Meehl, P. E. (1986). Trait language and behaviorese. In T. Thompson & M. D. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 315–334). Hillsdale, NJ: Erlbaum.

Michael, J., Palmer, D. C., & Sundberg, M. L. (2011). The multiple control of verbal behavior. *Analysis of Verbal Behavior, 27,* 3–22.

Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences, 7*(3), 141–144.

Ryan, J. J., Lopez, S. J., & Paolo, A. M. (1996). Digit span performance of persons 75–96 years of age: Base rates and associations with selected demographic variables. *Psychological Assessment*, *8*(3), 324–327. doi: http://dx.doi.org/10.1037/1040–3590.8.3.324

Schlinger, H. D., & Blakely, E. (1994). A descriptive taxonomy of environmental operations and its implications for behavior analysis. *Behavior Analyst, 17*(1), 43–57.

Sidman, M. (2010). Remarks. *Behavior and Philosophy, 38,* 129–132. (Original work published

1978)

Skinner, B. F. (1957). *Verbal behavior.* New York: Appleton-Century-Crofts.

Skinner, B. F. (1972). Current trends in experimental psychology. In B. F. Skinner (Ed.), *Cumulative record: A selection of papers* (pp. 223–241). New York: Appleton-Century-Crofts. (Original work published 1947) doi: http://psycnet.apa.org/doi/10.1037/11324–016

Staats, A. W. (1968). *Learning, language, and cognition.* New York: Holt, Rinehart and Winston.

Staats, A. W. (1996). *Behavior and personality: Psychological Behaviorism.* New York: Springer.