Processing limits of selective attention and working memory

Potential implications for interpreting

Nelson Cowan
University of Missouri

This article summarizes recent research on working memory and attention, with an emphasis on one theoretical framework in which working memory and attention are intricately related (Cowan, 1988, 1995, 1999a). Within this framework, working memory is conceived as an activated portion of long-term memory and, within that activated portion, the focus of attention and control processes that direct it. The focus of attention presumably can link activated elements to form new chunks of information. Several basic phenomena are sketched out, along with their potential relevance for the process of simultaneous interpreting. Current controversies that are likely to be of relevance to interpreting also are discussed. Although the question of how interpreters meet difficult processing demands cannot be answered presently because few studies have examined interpreting, the evidence does point the way to a number of promising lines of research on interpreting that could be carried out.

Processing limits of selective attention and working memory: Potential implications for interpreting

This chapter takes a look at research on attention and working memory in humans for the purpose of pointing out some methods, theoretical ideas, findings, and predictions from the field of cognitive psychology that may be relevant to the amazing task of simultaneously interpreting from one language to another. Interpreting involves the translation of spoken language as it is spoken or in what is termed "real time," which appears to require types of attention-sharing and overloading of working memory that people generally find very difficult as conditions of information processing.

Interpreting 5:2 (2000/01), **117–146.**
ISSN 1384–6647 / E-ISSN 1569–982X

Let us begin with a look at the origins and trends of research on attention and working memory and then examine some of the more recent research in these areas, with special emphasis on research issues that tie together attention and working memory. My own research has emphasized this relation between attention and working memory (cf. Cowan, 1988, 1995, 1999a, 2001). Avenues of research will be suggested that could address the role of attention and working memory in simultaneous interpreting.

Origins and trends of cognitive research on attention and working memory

The early days of information-processing research

There is a long history of philosophical thinking and research on attention and working memory, although relatively little application of this research to interpreting. In his text that can be taken as one of the most important precursors to cognitive psychology, William James (1890, pp. 403–404) asserted that "Every one knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought." In addition to providing some seminal ideas about attention, James also defined the concepts of "primary memory," referring to the very limited information that can reside in the conscious mind at any one time, and "secondary memory," referring to the vast amount of information that an individual retains over a lifetime. The concept of primary memory is related to attention if one makes the assumption that only attended events enter into primary memory. Roughly speaking, primary memory is the core of what is presently termed working memory (e.g., Baddeley, 1986), which can be defined as the limited amount of memory that can be used to hold information (and, according to some definitions, to manipulate it) as necessary for the execution of complex cognitive tasks.

Donald Broadbent and attention research

The modern success of research in these fields was greatly influenced by the technological advances that occurred during World War II and in the following decades and is still ongoing, including the development of general-purpose computers. Much of this research involved communication in ways that make it seem well-suited to the task of investigating the processes of interpreting (although it has not been applied to interpreting very often). Donald Broadbent summarized much of the research in his seminal 1958 book, *Perception and communication*. Psychologists in England contributed to the war effort partly by attempting to solve communications problems such as how multiple military airplanes could share the radio airwaves safely when a plethora of messages, including landing instructions, could be heard all at once. How well could a pilot single out the relevant message (i.e., the one addressed to him) from a sea of irrelevant messages? How well could this message be understood? How much could be understood as well from irrelevant messages? It turned out that this type of practical issue also had dramatic theoretical consequences, some of which I will discuss shortly.

Since the 1950s, there has been a great deal of new research clarifying processes of attention. (For a summary see Luck & Vecera, 2002.) However, the bulk of that research has used visual stimuli and therefore is not as closely relevant to the processes involved in interpreting as is the research on the processing of auditory stimuli including spoken language. The shift of the field from research primarily on auditory processes to visual processes of attention seems to have occurred because it became easier to achieve good stimulus control in the visual modality. The control of auditory stimuli requires, among other things, a relatively large amount of computer memory for the storage of speech digitized at rates that would be considered adequate for high-quality reproduction (e.g., storage of 15,000 voltage readings per second). With advances in personal computers since the 1980s, though, this sort of research has become feasible for more researchers. I will emphasize auditory research, which has continued as a stream aside the flood of visually-oriented research on attention since the 1970s.

George Miller, Alan Baddeley, and working-memory research

After James' (1890) concept of primary memory, the concept of working memory (though not yet by that name) received a boost from a widely-known article by Miller (1956) on "the magical number seven plus or minus two." This paper is related to the concept of attention in that, for primary memory as for attention, the central concept is that there is a limit to how much information can be held or processed at one time. As Miller pointed out in an autobiographical piece (Miller, 1989), he didn't take his own answer very seriously. He had two separate research areas that he needed to combine into an hour-long address: immediate recall of lists of various lengths, and absolute judgments of items on a continuum with a varying number of response choices. The only commonality he could find between the two research areas was a limit in the usual adult ability to about 7 items or response choices. He considered this possibly just a coincidence and his article focused instead on the process whereby items could be associated with one another to form a smaller number of coherent groups or "chunks" of information, and on how the limit in immediate recall seemed to apply to chunks rather than items. For example, it is difficult to remember 9 random letters but easy to remember letters forming 3 chunks, such as the 3 acronyms IBM – CIA – FBI, well known as organizations in the United States.

To my knowledge, the term "working memory" did not appear in James' writing. I believe that it also did not appear in the work of Broadbent (1958). Broadbent's maverick information-processing model was crammed into a footnote that referred to a "limited-capacity channel" to mean something like James' primary memory, whereas the vocabulary of his book in general was more closely tied to the behaviorist era. Atkinson & Shiffrin (1968), in a later version of an information-processing model in which rigorous mathematical modeling methods were applied, made popular the term "short-term memory" and placed emphasis on understanding control processes that shuttle information in and out of this short-term memory.

The term "working memory" may have started in animal research to mean memory for information that changed from trial to trial in the task, as opposed to other information that was constant throughout the task. However, the term was made more widely popular by an article by Baddeley and Hitch (1974) summarizing research on something like the primary memory system in humans. This research led to their suggestion that the concept of primary memory or short-term memory could reflect what happened in human experiments only if it were fractionated into several more special-purpose devices. They hinted at several devices that Baddeley (1986) later advocated more strongly: a phonological store that held verbal information; a visuo-spatial store that held information about the layout of objects in space; and a set of attention-demanding, central executive processes that were used to transform information, and to initiate processes (such as covert verbal rehearsal) that were then used to maintain information in storage. The research showed that verbal and spatial information interfered with like information much more than they interfered across different types of information. Thus, there was little interference between verbal and spatial information in memory.

Differences between several approaches

One difference between the frameworks offered by Broadbent (1958) versus Baddeley (1986) is that a different level of coding was emphasized. Broadbent's model placed a fair amount of emphasis on sensory memory codes. These were seen as relatively unprocessed and large in capacity (though short-lasting) but it was thought that we could draw only a small amount of information from sensory memory into limited-capacity processing. Baddeley and Hitch placed more emphasis on types of memory that were more processed, yet still passively held (and also short-lasting). For example, given that memory for printed letters is impaired when the letters have names that sound similar (B, D, T, V, P, etc.), it seemed that participants must generate a phonological code and that this internal phonological code was more important in working memory than was the peripheral sensory code.

Broadbent's (1958, p. 299) model did term the storage device that was limited by time, but not by capacity, a "short-term store." However, earlier chapters of his book make it seem that this included primarily the types of information that come from the senses. His model did allow that the limited-capacity system could feed information back to the short-term store but he did not clarify whether he thought that it was sensory information or more abstract coded information that was placed back into the short-term store. It has typically been assumed that only abstract information can be fed back to short-term storage or "rehearsed" but research by Keller, Cowan, and Saults (1995) shows that attention to the pitch of a tone during a 10-second retention interval can help to prolong information about that tone pitch with more precision than a semitone, suggesting that, in principle, either sensory or more abstract information, such as phonemes, could be rehearsed. Perhaps that was Broadbent's guess also.

Although it was reasonable for Baddeley (1986) to place great emphasis on short-term storage of abstract codes such as phonemes, research on a less-abstract sensory memory makes it clear that it, too, provides a rich source of information for a short time following each stimulus onset. Providing a partial-report cue shortly after an array of items is presented allows excellent recovery of the information in the array, both in the visual modality (Sperling, 1960) and in the auditory modality (Darwin, Turvey, & Crowder, 1972; Rostron, 1974). Yet, whole report of the array allows only about 4 items to be recovered. The large amount presumably comes from a relatively unprocessed sensory memory, whereas the limited amount comes from a form of memory that may be more consistently accessible to consciousness, akin to primary memory. The working memory model (Baddeley, 1986) has never quite integrated sensory memory into the model in a general sense. To take two examples, it says little or nothing about how tactile sensory information is encoded and it must use the phonological store as the device to save tone information.

The phonological memory that Baddeley and Hitch discussed (see also Baddeley, 1986) was said to be limited to a few seconds for a particular entry unless it is refreshed by rehearsal. This suggestion was made largely because individuals could recall about as many items as they could repeat in about 2 seconds, which could reflect the operation of a covert rehearsal loop. If an item was not refreshed before it decayed from memory, it would be lost. Overt and covert speech appear to have comparable speed limits within an individual (Landauer, 1962). Therefore, the rate of speech was taken to approximate the rate at which a series of items could be rehearsed and the 2-second limit was taken as the length of items that could be rehearsed in a repeating loop without being lost to decay.

It seems to have been assumed (at least implicitly) in Baddeley's model that phonological storage is limited to a single stream of speech at one time. This limit seems reasonable given that phonological analysis seems to take some attention and therefore would be difficult to carry out very well for multiple speech streams at the same time. An experiment by Cowan, Lichty, & Grove (1990, Experiment 4) demonstrates this. Their subjects carried out a task in which they were to read in a whisper while syllables were presented to them through earphones. Occasionally, reading was interrupted by a light cue and memory for the identity of the most recent syllable, which had just ended, was tested. Subjects' perception of the consonants within ignored syllables was much worse on trials in which the subjects read flawlessly than on trials in which their attention wandered from the whispered reading, in which case attention presumably went to the speech that was to be ignored.

In my own writing, I have respected the specificity of interference in working memory. For example, I have made the assumption that phonological material interferes with phonological codes from previous stimuli already in memory, spatial material interferes with spatial codes from previous stimuli, and so on. However, I have been uncomfortable accepting the generality of a working-memory system like Baddeley's that seems to reify a few modular, passive storage systems (the phonological loop and visuospatial sketchpad). Where, for instance, would tactile or nonverbal auditory information go in such a model? Instead, I have simply referred to sensory, phonological, visual, and other sorts of information together as activated subsets of long-term memory. I also have emphasized that a subset of the activated information receives a more complete analysis and is represented in the focus of attention, which also can include new links between elements (e.g., Cowan, 1995, 1999a; see also Massaro, 1975). Baddeley (2000, 2001) also has supplemented his system recently with an episodic buffer, which stores links between materials that do not fit cleanly into either the phonological buffer or the visuospatial buffer. This episodic buffer appears to serve much the same purpose as the focus of attention in my own model.

In order to illustrate the modeling possibilities, the Broadbent, Baddeley, and Cowan modeling frameworks are depicted in Figures 1 through 3, respectively. It should be noted that all three of these are rather rough modeling frameworks and, at this point, it is sometimes difficult to derive exact predictions from them for some experimental situations. However, I consider them helpful in conceiving of the entire information processing system and therefore organizing the concepts that can be used to motivate further research.

Research on the deployment of attention and working memory in simultaneous interpretation

There has been relatively little research tying attention and working memory to the process of interpreting. One elegant study in which this was done is a study by Gerver (1974). Experienced interpreters heard French prose passages and, in different conditions, were supposed to do one of the following: (1) just listen to the passage, (2) simultaneously interpret from French into English, or (3) shadow the prose. Shadowing is a task that was developed in the early days of cognitive psychology (see Broadbent, 1958) as a way to ensure that a passage was constantly attended, typically to divert attention from another stimulus channel or stream. After one of these three tasks, the participant would be asked to answer comprehension questions regarding the passage.

One possible expectation would be that interpreting would require the most attention and effort and that it therefore would not allow enough devotion of attention for the prose to be understood at a deep level. However, a contrasting expectation was that interpreting required a deeper level of encoding of the text than did shadowing, which could make for better retention of the meaning in interpreting. That was indeed the outcome. The percentage of correct responses on comprehension questions was 58% for simple listening, 51% for interpreting, and 43% for shadowing. Thus, simultaneous interpreting causes some distraction from the comprehension of meaning compared to simply listening, but that distraction seems relatively minor.

Moser-Mercer, Frauenfelder, Casado, and Künzli (2000) reported some more recent studies of what interpreters can and cannot do. It shows, for example, that professional interpreters are poorer at shadowing than student interpreters. Apparently, the expertise in interpreting is misused in this case so that shadowing is treated to some extent as an interpreting task, with changes in wording that are not appropriate to the assigned task. Professional interpreters also made fewer errors when reading with delayed auditory feedback, a situation that ordinarily is quite disruptive. All of this shows that their pattern of learning may help them to ignore distractions that commonly occur in an interpreting situation.

This research on interpreters does not bring us very far toward understanding what goes on in the brain during interpreting. There is basically an apparent contradiction between what happens in ordinary situations to typical individuals and what happens in interpreting. Cherry (1953) presented (to typical individuals) dichotic stimulation with one speech channel in one ear and a different channel in the other ear. Participants generally could concentrate on the contents of one ear but, in order to understand this one message, they were unable to remember much at all of what was presented in the channel that was to be ignored. They typically were aware of only the physical characteristics of the voice in the ignored channel and not the meaning.

Yet, we also know that it is possible to learn to do more. In an early study, Solomons and Stein (1896; this is the famous poet, Gertrude Stein) showed that they could learn to read to each other while taking dictation from the other person's reading. (For similar, more recent evidence see Hirst, Spelke, Reaves, Caharack, & Neisser, 1980.) The task in interpreting is not so dissimilar. The question, then, is how human capabilities can typically be so limited and yet allow the learning of special feats such as reading while taking dictation or simultaneous interpreting, which involves listening while translating. By emphasizing some of the areas of research on attention and working memory, I hope to shed some light on this question.

The relevance of research on various aspects of attention and working memory

In the following sections, issues relating attention and working memory will be examined and then, in each case, extensions to the investigation of simultaneous interpretation will be suggested. The issues stem from the various modeling frameworks described above. They include (1) the properties of attentional filtering, (2) the properties of inhibition and attentional control, (3) the speed with which information is retrieved into the focus of attention, (4) the capacity of the focus of attention, and (5) the rapid forgetting of unattended information in memory. In each case I will present the basic phenomenon, its relevance to interpreting, and current controversies and their relevance, in turn.

Attentional filtering

The basic phenomenon. Consider the situation in which a person is to shadow information from one channel, with a different message presented in the other channel that need not be attended (Cherry, 1953; Wood & Cowan, 1995a). Given that people typically do not notice rather blatant changes in the unattended channel (e.g., a change from one language to another), Broadbent (1958) proposed that human information processing includes a filter that excludes information from all but one channel defined on the basis of physical features (see Figure 1).

Inasmuch as subjects do consistently notice basic physical changes in the unattended channel, though, such as a change from one voice to another, Cowan (1988, 1995, 1999a) emphasized that some processing of the unattended channel must occur, that is not represented in Broadbent's model. The type of model shown in Figure 3 accounts for that processing by eliminating the attentional filter. Instead, in this model all information is processed to some extent and activates at least some features in memory (such as features representing voice frequency and timbre), but only certain information recruits attention and ends up getting the more complete processing. That more complete processing is afforded only to information in the focus of attention. A change in physical properties of an ignored stimulus can recruit attention automatically, through what is called an orienting response, presumably resulting from a comparison of the neural model of recent stimulation with new stimulation (see Sokolov, 1963). Alternatively, central executive processes can be used to keep attention on a particular item voluntarily. Paying attention is

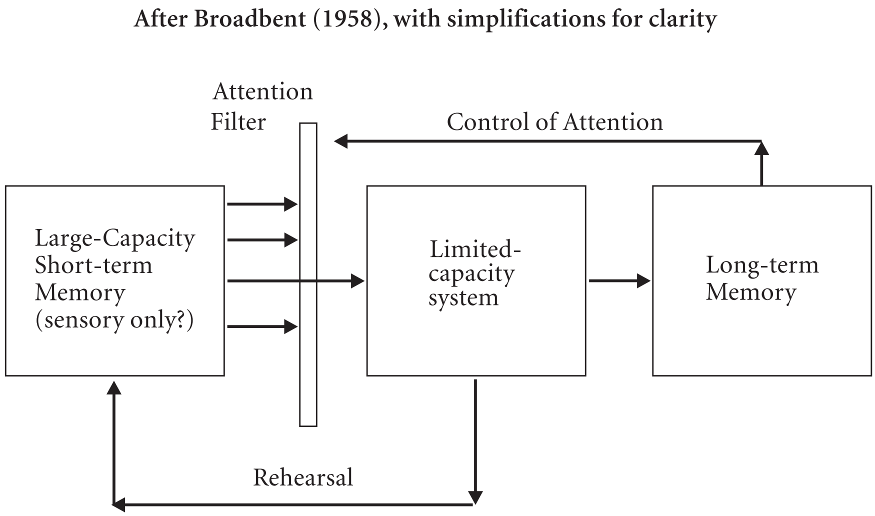

Figure 1. A depiction of the information-processing model of Broadbent (1958).

easiest when voluntary processes and orienting responses work in consonance, as when one wishes to pay attention to a lecturer who makes good use of changes in voice inflection and arm gestures to help hold one's attention. Other times, the two conflict, as when one listens to a lecture during a thunder-and-lightning storm (see Cowan, 1988, 1995).

Relevance to interpreting. The question for interpreting is how such a severe limit in attention under normal circumstances can be reconciled with what interpreters do. One could suggest that interpreters primarily listen to one channel and speak another, which may be different from listening to two channels at once. Moreover, the semantic content on the heard and spoken channels is supposed to be related, not unrelated. Yet, untrained individuals usually cannot do it well even if they know both languages well.

Theoretically, there are at least two possible ways in which interpreters may learn to overcome a limitation in attentional filtering. First, they may learn to switch attention rapidly and efficiently from what they are hearing to what they are saying and back again. This would allow them to maintain a maximum of one channel in the focus of attention at any moment. It can work because an automatic memory trace of the speech to be interpreted seems to linger in the mind in an acoustic or phonological form for a few seconds; for example, in a dichotic listening task in which one channel is shadowed, people can often report what is in the last second or so of the unattended channel after it ends abruptly (Broadbent, 1958; Glucksberg & Cowen, 1970; Norman, 1969). In interpreting, failing to attend to a short segment of speech would not be disastrous because attention could be turned to it slightly after the fact.

Second, it may be that the listening task or the speaking task becomes less attention-demanding after extended practice, in which case both tasks can occur at the same time. If we know which hypothesis is correct, that might have practical implications. If interpreting requires expert decisions regarding the switching of attention, that could be trained. If, on the other hand, expertise results in truly concurrent speaking and listening, then it would do little good to train attention-switching and emphasis should be placed more exclusively on practice and expertise. A detailed analysis of protocols from interpreters might be used to gain insight into the issue of attention-switching versus attention-sharing between listening and speaking. Also, interpreters could be tested on recognition of words that were spoken in the message to be interpreted while the interpreter was busy speaking concurrently, as opposed to words that were spoken while the interpreter was silent. An attention-switching hypothesis predicts that there will be somewhat better recognition of words presented while the interpreter was speaking.

Controversies regarding filtering. Whereas the earliest studies of filtering suggested that there is no processing of the unattended message, studies conducted soon afterward suggested that high-priority contents in the unattended channel somehow broke through the filter. For example, in an often-cited classic study, Moray (1959) carried out a dichotic listening experiment with shadowing of one channel and found that people sometimes noticed their own names spoken in the unattended channel in selective listening. This type of finding led Treisman (1964) to redefine filtering as an attenuation device that turns down or attenuates the unattended message without shutting it out entirely. Messages with higher priority become sufficiently activated to enter the focus of attention (cf. Norman, 1968). In Moray's study, one's own name would serve as a salient stimulus that would tend to be noticed. Some theorists even suggested that all stimuli are processed, the limit being only in how much of the information can be responded to at one time (e.g., Allport, 1980; Deutsch & Deutsch, 1963; Duncan, 1980). This became known as a "late filter" theory, whereas Broadbent's filter was an "early filter," filtering out stimuli from unattended channels before they had been processed much at all. The theoretical views of Treisman (1964), Norman (1968), and Cowan (1988) could be considered different types of "intermediate-filter" theories, suggesting that some, but not all, semantic information is extracted from unattended stimuli.

To the extent that semantic information actually is extracted from speech messages without attention, as the late-filter theory proposes, there is less of a riddle to be solved in order to understand how interpreters can listen and speak at the same time. Therefore, it is noteworthy that a recent study (Conway, Cowan, & Bunting, 2001) provides new support for Broadbent's early-filter theory. The background for the Conway et al. (2001) study was as follows. Moray (1959), though it has become an often-cited study, actually was little more than a pilot study finding that 4 of 12 subjects noticed their names in an unattended channel. Wood and Cowan (1995b) replicated this proportion (33%) almost exactly using more subjects and a better-controlled method. For example, subjects were yoked together in the experimental design so that two subjects (e.g., "Frank" and "John") both received both names in the unattended channel. Whereas subjects sometimes noticed their own names, they never seemed to notice the name of the yoked control subject. Noticing was judged not only by a post-experimental questionnaire, but also by hesitations and pauses in shadowing the relevant channel just after a name was presented. These on-line measures of attention shifting were highly correlated with the post-experimental questionnaire responses indicating that the name was noticed.

Nevertheless, we still must heed the cautions of Holender (1986), who pointed out potential problems in the control of attention in such studies. Perhaps the subjects who noticed their names did not do so automatically, as a late-filter theory would suggest, but noticed because their attention had not been focused fully upon the attended channel, as an early-filter theory would suggest. Conway et al. (2001) investigated this possibility by examining the relation between working-memory capacity and noticing one's name in Moray's procedure. Working-memory span was measured in a task in which subjects had to carry out mathematical operations while holding in mind an unrelated word presented after each operation. The working-memory span was the number of operations that could be carried out and the corresponding words correctly recalled afterward.

If the late-filter theory were correct, then all subjects would have a chance to notice their names. We know from Wood and Cowan (1995b), though, that this should not happen. It also might be the case that high-span individuals would notice their names more often because they would have enough attentional resources to shadow one channel and simultaneously monitor the other one. However, in contrast to this possibility, if the early-filter theory is correct, the only way that one could notice his or her name would be to fail to fix attention firmly on the shadowing task, allowing attention to wander to the channel that contained the name.

Conway et al. (2001) found that the early-filter interpretation clearly carried the day. Of the subjects in the lowest quartile of working memory, 65% recalled hearing their name, whereas, of subjects in the highest quartile of working memory, only 20% noticed their name. This suggests that noticing one's name did not come about through automatic semantic processing of both channels at once, but rather through a poorer control of attention in some subjects, leading attention sometimes to wander away from the assigned task and toward the other channel. The implication is that there is less processing of truly unattended information than was thought by Moray and by many others who have rejected the early-filtering view.

What of evidence that people can learn to do two similar things at once, such as reading and taking dictation (Hirst et al., 1980; Solomons & Stein, 1896)? Although this type of finding is fascinating without a doubt, Cowan (1995) argued at length that, in every such case, either some of the processing is becoming more automatic and less attention-demanding or subjects are learning to switch attention efficiently. It has long been clear that better memory retrieval allows procedures to be carried out more automatically, with less reliance on attention-demanding, effortful processes or algorithms (Logan, 1988; Shiffrin & Schneider, 1977).

A related hypothesis for interpreting is that, given human limitations, interpreters are unlikely to carry out their duties entirely by sharing attention between listening and speaking. If that is the case, we need to know the extent to which they improve by learning to switch attention rapidly and strategically between tasks, and the extent to which they improve by practicing until at least one of the tasks becomes much less attention-demanding. There might even be different styles of interpreting that differ in this regard.

Inhibition and attentional control

The basic phenomena. Many investigators have pointed to the importance of attention in inhibiting irrelevant information, maintaining focus on the goal of the assigned task, and managing to act in accord with the goal. The many relevant phenomena related to this function of attention have to do with correlations between working-memory tasks and the capabilities of attention.

This issue involving attentional control can be explained in terms of central executive processes. Baddeley (1986) discussed central executive processes as the collection of processes that plan, schedule, and execute various processes. However, much has been learned about the role of executive processes through

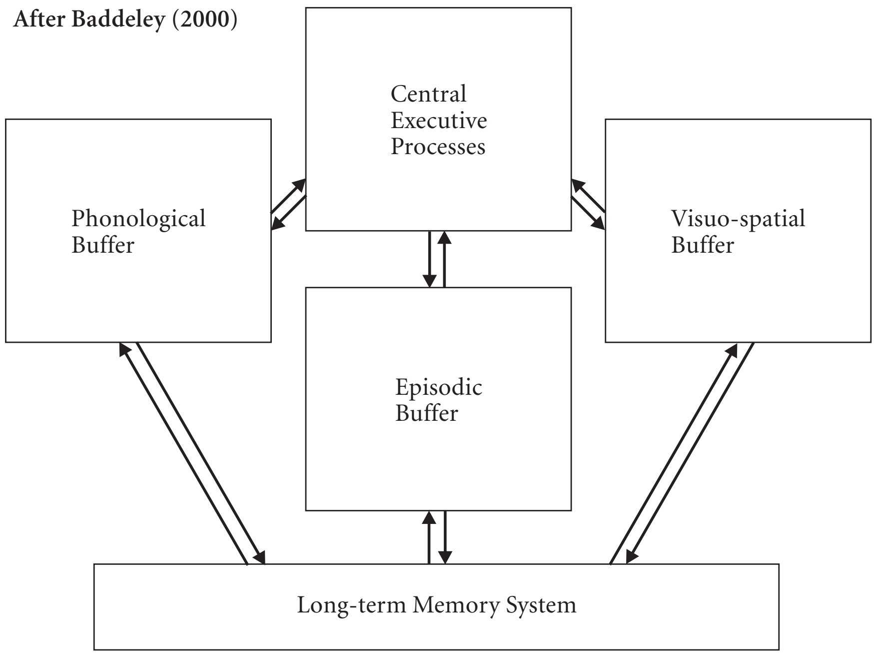

Figure 2. A depiction of the information-processing model of Baddeley (2000).

research centered around complex "working memory span" tasks, in which subjects carry out processing and storage concurrently. Daneman and Carpenter (1980) proposed such tasks on the grounds that both storage and processing must be engaged for the full capabilities of working memory to be tested. In their reading-span task, sentences were to be comprehended and the last word of each sentence was to be retained for later recall. The span was the number of sentences that could be processed and the sentence-final words recalled.

Such working-memory-span tasks, involving both storage and processing, are rather highly related to performance on intellectual aptitude and achievement tests; substantially higher than simple memory span (Daneman & Merikle, 1986; Mukunda & Hall, 1992). However, the reason for the high correlations is unclear. It may not be that they depend upon tasks that tie up both storage and processing, after all. It has been suggested that high-working-memory-span subjects excel at the ability to inhibit irrelevant information (Gernsbacher, 1993; Hasher, Stolzfus, Zacks, & Rypma, 1991) or, more generally, that they excel at the ability to control attention (Engle, Kane, & Tuholski, 1999; Engle, Tuholski, Laughlin, & Conway, 1999). For example, when one is reading, one often encounters words with more than one possible meaning, in a context that makes it possible for the reader to determine which meaning is correct. (An example is the word "bank," which has a different meaning in the contexts of "money" versus "river.") Good readers excel at holding in mind only the appropriate meaning and suppressing the inappropriate meaning, which presumably helps the reader to avoid distraction (Gernsbacher, 1993) and comprehend the intended message of the text.

Similarly, consider the situation in which color-words are presented in colors that sometimes conflict with the printed word (e.g., the word "blue" written in red ink) and sometimes do not conflict (e.g., the word "green" written in green ink). The task is to name the colors in which the word is presented. This is a version of the well-known Stroop task (Stroop, 1935) and it results in difficulty naming colors when they form conflicting color words. The more often the colors and words are consistent rather than in conflict, the harder it is to remember to ignore the printed words and name the colors. High working-memory span individuals are better able to stay on task in this situation, presumably because their attention is able to hold onto the goal better (Kane & Engle, in press).

Relevance to interpreting. A fundamental question about interpreting is whether anyone, given enough motivation and language expertise, could learn to do it or, if not, what fundamental individual abilities make someone potentially a good student of interpreting. Although many fundamental skills could be relevant, one of them is the control of attention. This skill would be especially important if it turns out that switching attention to and from the message to be interpreted at just the right moments is an important aspect of interpreting. It would also be important to the extent that any difficult task becomes more difficult as time goes on; vigilance is needed to stay on task. Perhaps some individuals would not be able to be good interpreters because their attention wanders after only a short period of very difficult concentration. Some of these skills may be inborn. At least, several recent studies suggest that performance on the complex types of working-memory task that appear to require the control of attention are closely related to the g factor in intelligence (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Engle et al., 1999).

In a cross-modal version of the Stroop task in which a color square was to be named as quickly as possible in the presence of an auditory distractor that could be a tone, a color-neutral adjective, or a conflicting color name (Elliott & Cowan, 2001), reaction times were slow at first and speeded up as the same distractor was repeated. They slowed down again when the distractor changed to a different word, especially a conflicting color word. Similarly, in the process of interpreting, there may be certain moments in which it is especially difficult to maintain adequate control, such as when the speech just has begun (Waters, McDonald, & Koresko, 1977), the speaker changes, or the topic changes.

Controversies regarding inhibition and attentional control. Earlier, I raised the question of whether interpreters carry out their ability primarily through good attention sharing between listening and speaking, or through strategic attention switching. The same controversy persists in the domain of working memory. When complex working-memory tasks first were developed, they were developed on the grounds that the working-memory system included both storage and processing components and therefore that both sorts of components had to be engaged at once in order to assess the true capability of working memory (Daneman & Carpenter, 1980). Such tasks correlated with scholastic and intellectual aptitude tasks better than simple span tests, such as the digit-span test that appears in intelligence test batteries. The assumption was that attention had to be shared between storage and processing components of the task. However, there are forms of working-memory task that do not include explicit, separate storage and processing components and, yet, still seem to provide equally good correlations with aptitude tests. For example, in running memory span, a long list of items (usually digits) is presented quickly and stops at an unpredictable point, after which the subject is to recall as many items as possible, in order, from the end of the list. Mukunda and Hall (1992) noted correlations between running memory span and intellectual abilities that were just as high as those between the storage-plus-processing type of task and intellectual abilities. Apparently, attention-sharing is not an essential aspect of working memory and may not be an essential aspect of attentional control capabilities.

Towse, Hitch, and Hutton (2000) and Hitch, Towse, and Hutton (2001) presented evidence that subjects in working memory span tasks often switch attention between processing and storage rather than sharing attention, and that the more successful subjects may be those who can complete the task more rapidly. Cowan et al. (in press) presented additional evidence favoring attention-switching in reading and listening span tasks. When response times were measured, they were much longer for these tasks than for a digit-span task or a counting-span task, in which screens of dots had to be counted and the sums remembered. Presumably, in reading or listening span, the subjects could recall each sentence as a cue to recalling the sentence-final word (a time-consuming process), whereas no such context was available in the digit- or counting-span tasks. Thus, strategic attention switching seems important to success in working-memory tasks, whereas the involvement of attention sharing perhaps is less clear.

Attentional control is not the only important skill that is involved in working memory. Conway et al. (2002) found that complex working-memory tasks did correlate with intelligence but they found a second, separate constellation of factors including memory for simple lists of verbal stimuli and processing speed. It is quite possible that interpreters must be good in both ways: in general intelligence, but also in verbal fluency and processing speed. One good way to begin to find out would be to test interpreters of various levels of achievement on the battery that Conway et al. (2002) used, including speed-of-processing tests, simple span tests, storage-plus-processing span tests, and intelligence tests.

Speed of retrieval into the focus of attention

The basic phenomenon. Some theorists believe that the difference between superior and inferior information processors is largely in the speed of processing (e.g., Fry & Hale, 1996; Kail & Salthouse, 1994; Salthouse, 1996). According to the theoretical framework of Cowan discussed above, this retrieval speed probably reflects the speed with which information can be transferred from memory into the focus of attention (see Figure 3).

One type of retrieval speed that may be of particular importance is the speed of retrieval in various sorts of working-memory tasks. These can be measured by using a computer to measure the time between the end of a spoken list and the beginning of a spoken recall response, the duration of every word in recall, and the duration of pauses between words. Within correctly-recalled lists, pauses between words in the response are longer for longer lists (e.g., Cowan et al., 1998; Tehan & Lalor, 2000). Therefore, it is clear that what is being processed during the pauses is list-wide rather than being restricted to planning of the output of the word that is about to be recalled. These pauses have been taken to reflect a process whereby the entire list is searched to determine which word to recall next (Cowan, 1992).

This memory-search process during serial recall appears to be based on abstract lexical representations rather than speech representations, inasmuch as the pauses do not depend on word length (Cowan et al., 1994) and are uncorrelated with rapid-rehearsal rates (Cowan et al., 1998). They depend heavily, though, on the lexical versus nonlexical status of verbal items (Hulme, Newton, Cowan, Stuart, & Brown, 1999). They may reflect the retrieval of verbal items from an activated form (e.g., the phonological buffer of Baddeley, 1986) into the focus of attention for recall, with the assistance of lexical knowledge.

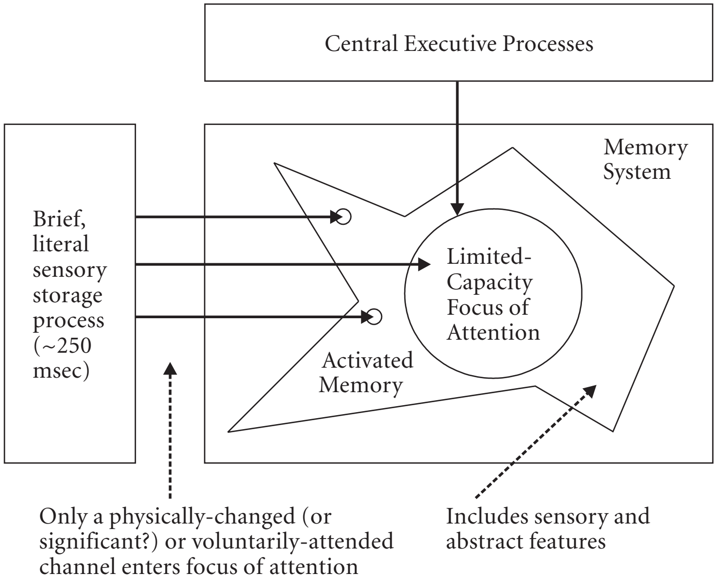

Figure 3. A depiction of the information-processing model of Cowan (1988).

Relevance to interpreting. Clearly, the faster comprehension, response formulation, and other processes are, the easier it should be to coordinate these processes in interpreting. The more we learn about retrieval speeds and what processes determine them, the more clues we will have about how to pinpoint the processes contributing to excellent interpreting and to learn how training might be improved.

Controversies regarding retrieval rates. A controversy here is that it is not clear how many separate retrieval rates there actually are. One hypothesis has been that there is a single, global processing speed in an individual that determines processing speeds on various tasks (e.g., Kail & Salthouse, 1994). However, Cowan et al. (1998) and Cowan (1999b) found that covert verbal rehearsal speed (measured by the ability to repeat items rapidly) and short-term memory retrieval speed (measured by the duration of pauses between words within serial recall responses) were unrelated to one another, and that they improved at very different periods of childhood development. This seems to implicate multiple sources of speed rather than a single, underlying global speed of processing.

Basically, it is unclear whether speed of processing is an explanation, stemming from some sort of neural integrity or efficiency, or whether it is merely a consequence of the improvement in some set of processes or in knowledge. To illustrate the two different ways in which speed can be interpreted, imagine that two elderly people and one young adult are all trying to be the first to find a missing television remote controller. Further, imagine that one of the elderly people was the last to use the controller. On one hand, the elderly people move slower than the young adult. On the other hand, one elderly person has some knowledge of where the controller was when it was lost. Therefore, two of the three people have an advantage over the remaining elderly person.

Translate this analogy into interpreting. An individual who naturally has an especially fast verbal fluency has an advantage in learning to carry out interpreting quickly, but so does an individual who has an outstanding knowledge of the two languages and an understanding of language principles generally. Both advantages will speed up interpreting performance, but detailed testing would have to be done before one would know what profile of abilities leads to the faster interpreting speed.

Capacity of the focus of attention

Basic phenomenon. Miller (1956) suggested that people can recall about 7 items but this estimate was admittedly rough and did not take into account the various ways in which people can group information. In fact, a main point of this seminal article was the incredible power of chunking, or grouping items together to form larger ones. In order to provide a better estimate of the capacity of memory, Cowan (2001) examined many different kinds of procedures in which it seemed reasonable to assume (1) that the presented items were very familiar, single chunks and (2) that the test circumstances prevented further grouping of items into larger chunks. In such circumstances, a uniform average limit of about 4 items was observed across various tests. It was suggested that this may reflect the limit in the capacity of the focus of attention (see Figure 3) in chunks. The important thing about this finding was not the particular estimate that was obtained but, rather, the prospect that it should be possible to find an estimate of capacity at all, whereas that had seemed beyond reach. Hopefully, the number of tasks will grow for which the size of chunks in the mental representation will become clear and we can see if the same capacity estimate applies across the board.

Relevance to interpreting. There is no question that an interpreter must hold in mind several types of information at once. There is at least one chunk of information, and probably more, that makes up the current thought to be translated. There is more information making up the manner in which the translated sentence was started so that the remainder of the sentence can be added in a grammatically consistent manner. There may be additional information regarding the nature of the audience or the occasion that has to be held in mind, as well as information about the larger context of what the speaker has been saying, which may influence the way in which the ongoing speech is interpreted. If people differ in the capacity of the focus of attention, they may differ accordingly in their potential to become interpreters. Also, an analysis of a specific interpreting situation (comprising the larger context as well as the immediate speech to be interpreted) in terms of the number of chunks that have to be held in mind at once may eventually allow predictions about what makes one interpreting situation more difficult than another.

One hypothesis of Cowan (2001) that seems especially relevant to a task such as interpreting is that separate items or chunks can be combined into a single, larger chunk only if they can be present in the focus of attention at the same time. In ordinary language comprehension, people may not always achieve that to a sufficient extent and consequently do not always achieve an adequate, consistent representation of what was spoken. In fact, Ferreira, Bailey, and Ferraro (2002) make the case that there are capacity limits in language comprehension that are more extreme than is usually realized. As a result, subjects often do not form a complete and consistent representation of every sentence. An example occurs for garden-path sentences such as: "While Anna dressed the baby played in the crib" (presented without commas). This sentence is difficult because, immediately after reading of the baby, readers often assume that it is the baby that Anna dressed. After finishing the sentence it should become clear that, actually, Anna dressed herself while the baby played. However, readers often will answer "yes" to the question of whether Anna dressed the baby, even while answering "yes" also to the question of whether the baby played. This is possible only if the sentence is encoded as a series of unconnected fragments or propositions, and not on the basis of any consistent syntactic interpretation of the sentence. The cause of this fragmented interpretation may be the unavailability of, or failure to use, sufficent attention to construct the correct interpretation and to suppress any incorrect interpretations.

A partial interpretation of the sentence often is enough for adequate understanding in casual conversation. In interpreting, though, that kind of partial sentence comprehension should not allow ideal performance. Failure to understand the full, correct meaning of one sentence should sometimes make for a poor translation of the next sentence. It seems as if it would be fruitful to examine how the level of an interpreter's comprehension of what has been spoken is related to the quality of the interpretation that is accomplished, and to see how often faulty interpretations can be attributed to misunderstandings of the preceding text. Moreover, it seems likely that lower performance on working-memory tasks would be related to comprehension errors of the type that Ferriera et al. (2002) discussed, and also to poorer interpreting abilities.

Controversies regarding the capacity of the focus of attention. Controversy is, at this point, very much present in the ideas regarding the capacity of the focus of attention, as is indicated by the many commentaries that are part of the Cowan (2001) article, and the reply to the commentaries. Not everyone agrees that the consistency in the observed capacity limit between the many studies reviewed by Cowan (2001) is more than coincidence. There are also recent studies in which authors have suggested that the focus of attention holds only one chunk at a time, not up to four or so (McElree, 2001; Oberauer, 2002). I believe that, although these studies have pointed out some important regularities of learning and memory, there are other explanations for their findings and they do not truly indicate that the focus of attention is limited to a single item. In any case, the evidence discussed by Oberauer (2002) still does suggest that there is a working-memory capacity limit of about 4 chunks, peripheral to the focus of attention in his opinion. If this is the case then the combination of activated elements to form new chunks must take place outside of the focus of attention per se. The implications for information processing would be similar to what Cowan (2001) suggested except that individual differences in capacity limits might not be expected to be related to individual differences in the management of the focus of attention.

Arguing against the notion that the capacity limitation is unrelated to attention processes, the running memory span task, which correlates well with intellectual tasks (Mukunda & Hall, 1992), is one of the tasks that yields an estimate of about 4 chunks. According to the analysis of Cowan (2001), the unknown list length and fast presentation rate prevent subjects from being able to form multi-item chunks in this task, so that information has to be retrieved from the perceptual memory stream into the focus of attention just after the list ends. At this point, each item is still a separate chunk and only about 4 items can be retrieved. Ongoing studies in our laboratory indicate that measures of the capacity of the focus of attention are, in fact, related to storage-plus-processing types of measures of working memory. This seems most consistent with the notion that attentional control and the capacity of the focus of attention are two closely related concepts. Individuals who are better at tasks measuring one of these should be better at both.

Of those who do agree that there is a regularity in the capacity limit to about 4 items, there are disagreements regarding its cause. It is theoretically possible that some principle of item distinctiveness limits the set of representations in any domain to about 4 and that this limit does not have to do with the focus of attention after all. We are approaching the problem currently by looking for capacity limits among sets of items that include very different members. For example, given that subjects can retain about 4 colors squares from an array or about 4 tones from a sequence, what can be retained if the two are presented concurrently? Can subjects retain a total of 8 items (e.g., 4 tones plus 4 color squares), or only a total of 4 as would be expected if retention occurs in the focus of attention? Is the answer the same no matter whether the stimuli are presented as arrays or as sequences? Our work currently focuses on this issue.

Unresolved issues such as these should have implications for interpreting. For example, when we understand better the nature of capacity limits we will understand the degree to which there are separate limits versus a combined limit in memory for several things: the target speech to be interpreted; memory for the interpreter's own recent responses, which must be retained by the interpreter for a while to achieve coherence; and any written notes or other relevant information.

Rapid forgetting of unattended memory

The basic phenomenon. There are many types of memory code that may survive for a while outside of the focus of attention, before being degraded by decay or interference (Cowan, 1995). For speech signals, this might include memory for the way in which the voice sounded, as well as some of the phonemic information. At least for speech signals that have been attended recently (but are no longer in the focus of attention), it could include semantic information also. For example, Cowan et al. (1990) presented syllables of speech through headphones at irregular intervals (several seconds apart) while subjects were busy reading a novel (silently or in a whisper), interrupting the reading occasionally to test for memory for the most recent syllable, which had been presented 1, 5, or 10 seconds ago. Memory was lost quickly across 10 seconds when the syllables were ignored in this way but, in several studies, when attention was divided between reading and speech or shifted entirely to speech, there was no such forgetting across 10 seconds.

Baddeley (1986) assumed that the entry of information into a phonological form was automatic and that the decay rate of phonological memory did not differ in an important manner among individuals. Although this has never been tested directly, a recent childhood developmental study of Cowan, Nugent, Elliott, and Saults (2000) provides some insight. The study used a silent, primary task involving matching pictures based on rhymes among their names. Lists of digits of a length equal to the subject's span were presented in headphones and were to be ignored. Occasionally, the rhyming task was interrupted and memory for the last digit list was tested. Averaged across all serial positions of the list, there was no difference between second-grade children, fifth-grade children, and adults in the observed rate of decay over a 10-second retention interval. This seems to support Baddeley's assumption of no individual differences in the rate of loss of memory for unattended information.

However, Cowan et al. (2000) did find that, for the final serial position observed separately, there was a striking difference in the rate of loss in the younger versus the older children. With a 1-second retention interval they were both at about the same level (68% vs. 72% correct) but, after a 5-second interval, the second-grade children had dropped to only 36% correct, whereas the fifth-grade children were still performing at 63% correct. This suggests that a specific type of information does persist longer in older children. It is a type of information that is unique to the final list item and therefore cannot withstand interference from subsequent item presentations. This would appear to be auditory sensory memory, or memory for the acoustic qualities of the item (Cowan, 1994).

Even in adults, whereas memory for a speech item persists for at least 20 seconds, memory for its acoustic quality is much shorter-lived. Balota and Duchek (1986) found this by presenting lists followed by an interfering final item that did not have to be recalled, or "suffix," followed by serial recall. A suffix typically interferes with memory for items toward the end of the list. What was unique about their experiment was the presentation of a suffix either immediately or after a 20-second interval (filled with a distracting task). On trials in which the suffix was immediate, a speech suffix interfered with recall more than a tone suffix and a suffix in the same voice as the list interfered with recall more than a different-voice suffix. On trials in which the suffix was delayed by 20 seconds, the speech suffix still interfered more than the tone but the difference between same- and different-voice suffixes no longer mattered. Presumably, the memory that persisted to be interfered with included detailed acoustic information at first, but not after a 20-second delay.

Relevance to interpreting. Inasmuch as the interpreter has multiple tasks to accomplish at once (i.e., listening, formulating a translation, and speaking), interpreting might be possible only because some automatic forms of memory, such as memory for ignored speech, relieve some of the demand placed upon attentional mechanisms. For example, an interpreter might be able to withdraw attention from the speech that is to be interpreted for a second or so while thinking of a response to the previous phrase, because the unattended speech will still be represented in auditory sensory and/or phonological memory and can be attended to at that slightly later point in the process.

Certain voices or accents might leave a better speech memory representation than others, and certain individuals may have better auditory memory retention than others. All of this is worth systematic investigation.

Controversies regarding rapid forgetting of unattended memory. Causing of a shift of attention away from speech and then testing for memory of that speech is a method that some investigators have used to establish that the memory is automatically held for a while. Another method to investigate automatic processes is to examine whether speech that is irrelevant to the ongoing task can be ignored or whether it interferes with processing of a printed list of words. At first glance, these two methods have yielded what appear to be contradictory results. Cowan et al. (1990) found that memory for ignored phonemes was impoverished, whereas studies of the effects of irrelevant speech show that they interfere with memory for printed lists and thus must be encoded in some way despite the subject's intentions (e.g., Salamé & Baddeley, 1982).

The apparent contradiction may be resolved by the finding that the interference in irrelevant-speech procedures does not depend on the phonemic similarity between the irrelevant speech and the printed list items (Jones & Macken, 1995; LeCompte & Shaibe, 1997). It may be that acoustic, low-level aspects of irrelevant sounds are represented automatically and inevitably interfere with memory for printed words, which explains why irrelevant tones also interfere with this memory quite a bit (Jones & Macken, 1993). In contrast, phonological encoding may be assisted by attention (Cowan et al., 1990) and therefore the details of the phonological representation of the irrelevant speech, which the subject tries to ignore, would not be so important in determining the amount of interference.

Finally, Baddeley, Gathercole, and Papagno (1998) summarized considerable evidence that the integrity of phonological memory storage is important for vocabulary learning. For example, children who can repeat longer nonwords tend to learn more vocabulary. However, many of the parameters of this phonological memory storage have yet to be examined. Does better storage entail a higher capacity, a more precise representation of speech, a longer-lasting representation, or what exactly? The theoretical framework that I have adopted (Cowan, 1995) would suggest that there is no capacity limit on this automatic sort of memory but that its precision and persistence could differ among individuals.

There are a number of cognitive procedures in this arena that would be interesting to examine in interpreters in comparison with non-interpreters, and in expert versus novice interpreters. Topics could include the persistence of memory for ignored speech sounds and the effects of irrelevant speech on performance. The basic strategy for such research is to try to find tasks for which any differences between interpreters and other people could not easily be attributed to training in interpreting and therefore would have to be attributed to differences in the kinds of individuals who succeed at entering the profession. One thing that would be very helpful would be a longitudinal study following individuals from the beginning of their training in interpreting to a later date at which some have become proficient and others have not.

Concluding remarks

In this paper I have pointed out novel measures for a variety of basic processes in the areas of working memory and attention. There are important limits in human processing related to the properties of attentional filtering, the properties of inhibition and attentional control, the speed with which information is retrieved into the focus of attention, the capacity of the focus of attention, and the rapid forgetting of unattended information in memory. At present, the research on these topics is mostly on normal individuals and therefore cannot resolve the question of what makes for an excellent interpreter as opposed to a mediocre or poor one. However, the research findings do suggest avenues of investigation likely to be fruitful, and they point to unresolved issues to be taken into account in the interpretation of research on interpreting.

References

Allport, D.A. (1980). Attention and performance. In G. Claxton (Ed.), *Cognitive psychology: New Directions* (pp. 112–153). London: Routledge & Kegan Paul.

Atkinson, R.C., & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence & J.T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). New York: Academic Press.

Baddeley, A.D. (1986). *Working memory*. Oxford Psychology Series #11. Oxford: Clarendon Press.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in cognitive sciences, 4,* 417–423.

Baddeley, A. (2001). The magic number and the episodic buffer. *Behavioral and Brain Sciences, 24*, 117–118.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158–173.

Baddeley, A., & Hitch, G.J. (1974). Working memory. In G. Bower (ed.), *Recent advances in learning and motivation,* Vol. VIII. New York: Academic Press.

Balota, D.A., & Duchek, J.M. (1986). Voice-specific information and the 20-second delayed suffix effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 509–516.

Broadbent, D.E. (1958). *Perception & communication.* London: Pergamon Press.

Cherry, E.C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America, 25*(5), 975–979.

Conway, R.A., Cowan, N., & Bunting, M.F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review, 8,* 331–335.

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,* 163–183.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin, 96*, 341–370.

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. (1992). Verbal memory span and the timing of spoken recall. *Journal of Memory and Language, 31,* 668–684.

Cowan, N. (1995). *Attention and memory: An integrated framework.* Oxford Psychology Series, No. 26. New York: Oxford University Press. (Paperback edition: 1997)

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

Cowan, N. (1999 b). The differential maturation of two processing rates related to digit span. *Journal of Experimental Child Psychology, 72*, 193–209.

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

Cowan, N., Keller, T., Hulme, C., Roodenrys, S., McDougall, S., & Rack, J. (1994). Verbal memory span in children: Speech timing clues to the mechanisms underlying age and word length effects. *Journal of Memory and Language, 33*, 234–250.

Cowan, N., Lichty, W., & Grove, T.R. (1990). Properties of memory for unattended spoken syllables. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16*, 258–269.

Cowan, N., Nugent, L.D., Elliott, E.M., & Saults, J.S. (2000). Persistence of memory for ignored lists of digits: Areas of developmental constancy and change. *Journal of Experimental Child Psychology, 76*, 151–172.

Cowan, N., Towse, J.N., Hamilton, Z., Saults, J.S., Elliott, E.M., Lacey, J.F., Moreno, M.V., & Hitch, G.J. (in press). Children's working-memory processes: A response-timing analysis. *Journal of Experimental Psychology: General.*

Cowan, N., Wood, N.L., Wood, P.K., Keller, T.A., Nugent, L.D., & Keller, C.V. (1998) . Two separate verbal processing rates contributing to short-term memory span. *Journal of Experimental Psychology: General, 127*, 141–160.

Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19*, 450–466.

Daneman, M., & Merikle, P.M. (1996). Working memory and language comprehension: A Meta-Analysis. *Psychonomic Bulletin & Review, 3*, 422–433.

Darwin, C.J., Turvey, M.T., & Crowder, R.G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology, 3*, 255–267.

Deutsch, J.A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70*(1), 80–90.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review, 87*(3), 272–300.

Elliott, E.M., & Cowan, N. (2001). Habituation to auditory distractors in a cross-modal, color-word interference task. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27*, 654–667.

Engle, R.W., Kane, M.J., & Tuholski, S.W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In A. Miyake & P. Shah (eds.), *Models of working memory: Mechanisms of active maintenance and executive control.* Cambridge, UK: Cambridge University Press. (Pp. 102–134)

Engle, R.W., Tuholski, S.W., Laughlin, J.E., & Conway, A.R.A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*, 309–331.

Ferreira, F., Bailey, K.G.D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*, 11–15.

Fry, A.F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science, 7*, 237–241.

Gernsbacher, M.A. (1993). Less skilled readers have less efficient suppression mechanisms. *Psychological Science, 4*, 294–298.

Gerver, D. (1974). Simultaneous listening and speaking and retention of prose. *Quarterly Journal of Experimental Psychology, 26*, 337–341.

Glucksberg, S., & Cowen, G.N., Jr. (1970). Memory for nonattended auditory material. *Cognitive Psychology, 1,* 149–156.

Hasher, L., Stolzfus, E.R., Zacks, R.T., & Rypma, B. (1991). Age and inhibition. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 17*, 163–169.

Hirst, W., Spelke, E.S., Reaves, C.C., Caharack, G., & Neisser, U. (1980). Dividing attention without alternation or automaticity. *Journal of Experimental Psychology: General, 109,* 98–117.

Hitch, G.J., Towse, J.N., & Hutton, U. (2001). What limits children's working memory span? Theoretical accounts and applications for scholastic development. *Journal of Experimental Psychology: General, 130*, 184–198.

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral & Brain Sciences, 9,* 1–66.

Hulme, C., Newton, P., Cowan, N., Stuart, G., & Brown, G. (1999). Think before you speak: pause, memory search and trace redintegration processes in verbal memory span. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 447–463.

James, W. (1890). *The principles of psychology*. NY: Henry Holt.

Jones, D.M., & Macken, W.J. (1993). Irrelevant tones produce an 'irrelevant speech effect': Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 19,* 369–381.

Jones, D.M., & Macken, W.J. (1995). Phonological similarity in the irrelevant speech effect: Within- or between-stream similarity? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 103–115.

Kail, R., & Salthouse, T.A. (1994). Processing speed as a mental capacity. *Acta Psychologica, 86,* 199–255.

Kane, M.J., & Engle, R.W. (in press). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General.*

Keller, T.A., Cowan, N., & Saults, J.S. (1995). Can auditory memory for tone pitch be rehearsed? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 635–645.

Landauer, T.K. (1962). Rate of implicit speech. *Perceptual & Motor Skills, 15,* 646.

LeCompte, D.C., & Shaibe, D.M. (1997). On the irrelevance of phonology to the irrelevant speech effect. *Quarterly Journal of Experimental Psychology, 50A,* 100–118.

Logan, G.D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

Luck, S.J., & Vecera, S.P. (2002). Attention. In Pashler, H., & Yantis, S. (Eds.), *Steven's handbook of experimental psychology* (pp. 235–286) (third edition). New York: Wiley.

Massaro, D.W. (1975). *Experimental psychology and information processing.* Chicago: Rand McNally.

Massaro, D.W., & Shlesinger, M. (1997). Information processing and a computational approach to the study of simultaneous interpretation. *Interpreting, 2,* 13–53.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 817–835.

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

Miller, G.A. (1989). George A. Miller. In L. Gardner (Ed.), *A history of psychology in autobiography, Vol. VIII.* (pp. 391–418). Stanford, CA: Stanford University Press.

Miyake, A., & Shah, P. (Eds.) (1999). *Models of Working Memory: Mechanisms of active maintenance and executive control.* Cambridge, U.K.: Cambridge University Press.

Moser-Mercer, B., Frauenfelder, U.H., Casado, B., & Künzli, A. (2000). Searching to define expertise in interpreting. In B.Englund Dimitrova & K. Hyltenstam (Eds.), *Language processing and simultaneous interpreting* (pp. 107–132). Amsterdam: John Benjamins.

Mukunda, K.V., & Hall, V.C. (1992). Does performance on memory for order correlate with performance on standardized measures of ability? A meta-analysis. *Intelligence, 16,* 81–97.

Norman, D.A. (1968). Toward a theory of memory and attention. *Psychological Review, 75*(6), 522–536.

Norman, D.A. (1969). Memory while shadowing. *Quarterly Journal of Experimental Psychology, 21*, 85–93.

Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 411–421.

Rostron, A.B. (1974). Brief auditory storage: Some further observations. *Acta Psychologica, 38,* 471–482.

Salamé, P., and Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior, 21*, 150–164.

Salthouse, T.A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103,* 403–428.

Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127–190.

Sokolov, E.N. (1963). *Perception and the conditioned reflex.* NY: Pergamon Press.

Solomons, L.M., & Stein, G. (1896). Normal motor automatism. *Psychological Review, 3*, 492–512.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74* (Whole No. 498.)

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. Journal of *Experimental Psychology, 18*, 643–662.

Tehan, G., & Lalor, D.M. (2000). Individual differences in memory span: The contribution of rehearsal, access to lexical memory, and output speed. *Quarterly Journal of Experimental Psychology, 53A,* 1012–1038.

Towse, J.N., Hitch, G.J., & Hutton, U. (2000). On the interpretation of working memory span in adults. *Memory & Cognition, 28,* 341–348.

Treisman, A.M. (1964b). Selective attention in man. *British Medical Bulletin, 20*, 12–16.

Waters, W.F., McDonald, D.G., & Koresko, R.L. (1977). Habituation of the orienting response: A gating mechanism subserving selective attention. *Psychophysiology, 14*(3), 228–236.

Wood, N., & Cowan, N. (1995a). The cocktail party phenomenon revisited: Attention and memory in the classic selective listening procedure of Cherry (1953). *Journal of Experimental Psychology: General, 124,* 243–262.

Wood, N., & Cowan, N. (1995b). The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 255–260.

Author's address

Nelson Cowan
Department of Psychological Sciences
University of Missouri
210 McAlester Hall
Columbia, Missouri, USA

E-mail: CowanN@missouri.edu

About the author

Nelson Cowan received his Ph.D. in Psychology from the University of Wisconsin, Madison, in 1980. After a postdoctoral fellowship at New York University and three years at the University of Massachusetts, Amherst, in 1985 he moved to the University of Missouri, Columbia, where he is currently Professor. He was associate editor of the *Journal of Experimental Psychology: Learning, Memory, and Cognition* (1995–2000) and is presently associate editor of the *Quarterly Journal of Experimental Psychology* (Section A). He wrote *Attention and memory: An integrated framework* (1995, Oxford University Press) and conducts research on working memory, short-term memory, and their relation to selective attention.