

## The Evolution of Language: The Cerebro-Cerebellar Blending of Visual-Spatial Working Memory with Vocalizations

Larry Vandervert

*American Nonlinear Systems*

Leiner, Leiner, and Dow proposed that the co-evolution of cerebral cortex and the cerebellum over the last million years gave rise to the unique cognitive capacities and language of humans. Following the findings of recent imaging studies by Imamizu and his colleagues, it is proposed that over the last million or so years language evolved from the *blending* of (1) decomposed/re-composed contexts or “moments” of visual-spatial experience with (2) those of sound patterns decomposed/re-composed from parallel context-appropriate vocalizations (calls or previously acquired “words”). It is further proposed that the adaptive value of this blending was the progressively rapid access to the control of detailed cause-and-effect relationships in working memory as it entered new and challenging environments. Employing the complex syntactical sequence of nut-cracking among capuchin monkeys it is proposed how cerebro-cerebellar blending of low-volume vocalization and visual-spatial working memory could have produced the beginnings of the phonological loop as proposed by Baddeley, Gathercole, and Papagno. It is concluded that the blending of cerebellar internal models in the cerebral cortex can explain the evolution of human advancements in the manipulation of cause-and-effect ideas in working memory, and, therefore, the emergence of the distinctive “cognitive niche” of humans proposed by Tooby and DeVore and supportively elaborated by Pinker.

Keywords: cerebellum, language evolution, visual-spatial working memory

Before humans evolved, the rate of enlargement of the basal ganglia exceeded that of the cerebellum, but when humans evolved, a change occurred: The basal ganglia enlarged at a particularly slow rate, while the cerebellum outstripped it, enlarging dramatically (Passingham, 1975; Stephan and Andy, 1969). It has often been remarked that an explanation is required for the threefold to fourfold increase in size of the cerebellum that occurred in the last million years of evolution (Washburn and Harding, 1970). [Leiner, Leiner, and Dow, 1986, pp. 443–444]

---

I thank Raymond Russ and an anonymous reviewer for comments on an earlier version of this article. Correspondence concerning this article should be addressed to Larry R. Vandervert, Ph.D., American Nonlinear Systems, 1529 W. Courtland Avenue, Spokane, Washington 99205. Email: lvandervert@aol.com

In their watershed article cited above, Leiner, Leiner, and Dow went on to propose that the greatly enlarged, phylogenetically newest parts of the human cerebellum (especially the ventrolateral portion of the dentate nucleus) might do for the rapid manipulation of *ideas* what the older parts had done for motor skills — namely, “serve as a fast information-processing adjunct of the association cortex” (1986, p. 444). In a follow-up article, Leiner, Leiner, and Dow (1989) extended their proposal to the combined mental and motor features of language: “two-way connections linking the cerebellum to Broca’s area (areas 44 and 45) in the inferior frontal convolution make it possible for the cerebellum to improve language dexterity, which combines motor and mental skills” (p. 1006). Thus, according to Leiner, Leiner, and Dow, the adaptive advantage of cerebro-cerebellar evolution in humans was the rapid and increased skill in the manipulation of ideas and progressive increases in language dexterity.

### *The Evolution of Language as an Adaptive Extension of Working Memory*

If existing non-human primates are any indication, all early hominins (notably *Homo habilis* a million and half years ago) had well-developed visual-spatial working memories. The visual-spatial working memories of monkeys have been found to be well-structured in spatial reasoning (Fragaszy and Cummins–Sebree, 2005) and in sequences of abstract reasoning (Obayashi et al., 2007). Therefore, as Aboitiz, Garcia, Bosman, and Brunetti (2006) and Vandervert (1997, 2003, 2009, in press) have suggested, the evolution of language would be most profitably studied as a direct adaptive extension of brain mechanisms which sub-serve visual-spatial working memory.

### *Purpose*

The purpose of this article is to describe the evolution of specific cerebro-cerebellar mechanisms which, it is hypothesized, resulted in the blending of visual-spatial working memory with vocalizations of early hominins to produce human language. Specifically, it is hypothesized that during approximately the last million years of cerebro-cerebellar *co-evolution*, language evolved from the cerebral *blending* of *multiple cerebellar internal models* of (1) decomposed/re-composed contexts or “moments” of visual-spatial experience *with* (2) those of sound patterns decomposed/re-composed from parallel context-appropriate vocalizations (calls or previously acquired “words”). It is proposed that the adaptive value of this blending was the progressively rapid access in working memory to the control of detailed cause-and-effect relationships in new and challenging environments. Before presenting further details of these mechanisms, it is necessary to describe relevant theoretical and empirical aspects of visual-spatial and speech-related working memory, the latter of which, according to Baddeley, Gathercole, and

Papagno (1998), evolved primarily to mediate the acquisition of new sound forms.

### *Working Memory*

Working memory has been described by Baddeley (1992) as a multi-component “brain system that provides temporary storage and manipulation for complex cognitive tasks such as language comprehension, learning, and reasoning” [abstract]. Baddeley divided working memory into the following three subcomponents: (1) an attention-controlling system which serves as a “central executive,” (2) a visual-spatial sketchpad which manipulates visual images within an ongoing flow of visual-spatial experience, and (3) a phonological loop which both *stores* and *rehearses* speech-based information. Baddeley, Gathercole, and Papagno (1998, p. 159) argued that the primary function of the phonological loop (both in silent and overt speech) is to learn the sound patterns of *new* words and new syntactical sequences and thereby to mediate language learning. In their conclusion, Baddeley et al. extended the phonological loop’s function of learning new sounds to the *evolution* of language: “the primary purpose for which the phonological loop evolved is to store unfamiliar sound patterns while more permanent memory records are being constructed [in long-term memory]” (1998, abstract). The storage and rehearsal process of the phonological loop involve the lateral cerebellum and speech-related areas of the cerebral cortex in both overt and silent speech used in solving problems (Marvel and Desmond, 2010, in press)

### *Details of the Cerebro-Cerebellar Blending Mechanism*

Imamizu, Kuroda, Miyauchi, Yoshioka, and Kawato (2003) and Imamizu et al. (2000) demonstrated the learning of multiple cognitive internal models in the lateral cerebellum (see Figure 1). Through the decomposition and re-combination of existing mental models, cerebellar internal models contribute to motor, sensory, and higher cognitive functions of the cerebral cortex in the optimization of goal-directed behavior (Imamizu and Kawato, 2009, in press). Imamizu, Higuchi, Toda, and Kawato (2007) found that when confronting *new* situations, these cognitive internal models were *blended* in the cerebral cortex to negotiate the new challenges. Based upon these findings they argued that cerebral blending of multiple cerebellar internal models bestowed several tightly inter-related advantages: (1) interference between different learning epochs [or “moments”] is reduced thereby enabling the rapid switching of skilled behaviors, (2) entirely new environments can be coped with by adaptively blending pre-existing motor and cognitive primitives as multiple internal models, (3) multiple internal models are blended *in proportion* to the requirements of the current new

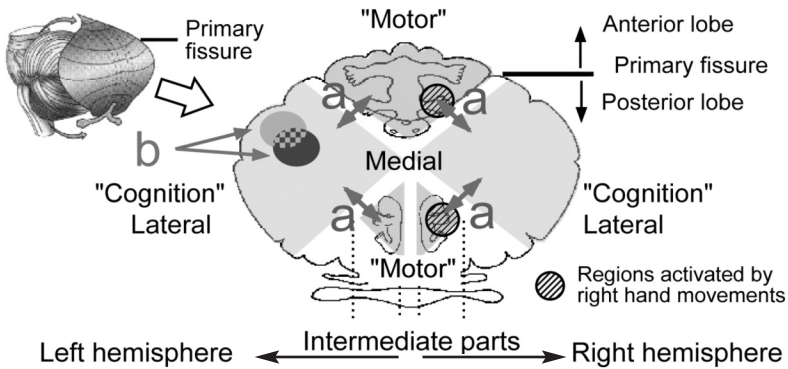


Figure 1: Flattened view of cerebellar surface illustrating that the anterior lobe and intermediate parts of the posterior lobe are related to “motor and somatosensory functions,” whereas the lateral posterior cerebellum is related to “cognitive functions.” To orient properly to the anterior/posterior axis of the flattened view, the viewer should keep in mind that anterior/posterior refer to what is actually a substantially convex cerebellar surface (see smaller drawing to left). Arrows at (a) indicate difference between “motor” (note modularity of somatotopic maps at top and bottom) and “cognition” found in previous neuroimaging studies. Arrows at (b) indicate modularity within the lateral posterior cerebellum for two different cognitive functions. From “Modular Organization of Internal Models of Tools in the Human Cerebellum,” by H. Imamizu, T. Kuroda, S. Miyauchi, T. Yoshioka, and M. Kawato, *Proceedings of the National Academy of Sciences*, 100, pp. 5461–5466. Copyright 2003 National Academy of Sciences, U.S.A. Reprinted with permission.

context, and (4) because blending is proportionate to the specific requirements of changing contexts, an enormous, perhaps limitless, repertoire of behavior can be generated even when the number of internal models might be limited. When applied to Baddeley, Gathercole, and Papagno’s (1998) conclusions on the evolutionary function of the phonological loop, this means that an enormous number of novel sound forms of new words could be mixed or blended with an equally enormous number of new visual-spatial contexts or moments.<sup>1</sup> It is proposed that the blending of multiple cerebellar internal models via the mechanisms found by Imamizu et al. (2007) can account for the evolution of working memory’s phonological loop and therefore the evolution of language.<sup>2</sup>

<sup>1</sup>Since, like words, particular sound patterns fractionated from vocalizations have no inherent meanings (no inherent, pre-determined connections with particular moments of visual-spatial experience) each group of early humans would have developed a slightly different “language” depending on the specific environmental circumstances and individuals involved.

<sup>2</sup>It is proposed that this feature of the nearly limitless blending of internal models of sound patterns and visual-spatial imagery explains the origin of what Hockett (1960) referred to as the “duality of patterning” feature of language (meaningless sounds or symbols can be rearranged to produce an unlimited number of messages, e.g., Hockett described how Morse Code exemplifies this feature). Hockett argued that duality of patterning is unique to human language. However, since monkeys have shown fronto-cerebellar action in switching tools (Obayashi et al., 2002)

*Cerebro-Cerebellar Decomposition/Re-Composition and Blending Within Working Memory*

Combining Baddeley, Gathercole, and Papagno (1998), Imamizu, Higuchi, Toda, and Kawato (2007), and Mandler (2004, in press) the following scenario for both the evolution of language in phylogeny and the development of language in ontogeny is proposed. In phylogeny, new environmental challenges set in motion the decomposition and re-composition of cerebellar internal models (Flanagan et al., 1999; Nakano et al., 2002) related to situation-specific visual-spatial moments and of likewise decomposed/re-composed vocalization patterns linked to those situation-specific moments. These new situation-specific visual-spatial moments and their linked situation-specific sound patterns were *blended* proportionately to meet the requirements of the new, challenging situation (Imamizu et al., 2007). It is proposed that these composites of blended cerebellar internal models were adaptive only because they represented *cause-and-effect meanings* in the new environmental situation.<sup>3</sup> It is further proposed that the adaptive value of this blending across generations of early humans was the progressively more rapid access in working memory to the control of progressively more detailed cause-and-effect relationships in the new, challenging environments. The blending process would have resulted in the gradual emergence of a working memory where moments representing cause-and-effect relationships could be quickly tagged into long-term memory using sub-vocal or vocal tags and which, subsequently, could be rapidly accessed from long-term memory using, again, sub-vocal or vocal tags to meet a variety of fast-moving environmental situations. The foregoing scenario provides an explanation for Baddeley et al.'s conclusion that the phonological loop selectively evolved due to its adaptive advantage of temporarily storing unfamiliar sound patterns while more permanent pathways were being constructed in long-term memory. The evolution of the phonological loop within the pre-existing visual-spatial working memory (and long-term memory) enabled the social sharing of detailed cause-and-effect relationships as well as the silent manipulation of ideas in planning, including the envisioning and manufacture of progressively advanced stone-tool technology.

---

indicating an open-ended synthesis of multiple *visual-spatial* internal models, duality of patterning appears to be shared, at least in nascent form, with other primate species, and, therefore, that duality patterning originates not in the tags that place moments of visual-spatial working memory in long-term memory, but in the limitless potential of internal models of those visual spatial moments themselves.

<sup>3</sup>Blended internal models are cause-and-effect models simply because in Imamizu et al.'s (2000, 2003, 2007) imaging studies their acquisition leads subjects to skillful cognitive and motor manipulation of new tools. In evolution, then, the blending of internal models was adaptive, because it bestowed the advantage of the skillful manipulation of cause-and-effect relationships (threatening or opportunistic) in the environment.

In ontogeny, it is proposed that the foregoing blending of visual-spatial “moments” and linked sound patterns account for the formation of conceptual primitives and subsequent language acquisition (including the emergence of the phonological loop) in infants. The above account of the decomposition/re-composition of visual-spatial experience and its blending with situation-specific sound patterns provides a supportive neurological basis for Mandler’s (1992a, 1992b, 2004, in press) theory of early conceptual and language development involving how perceptual meaning analysis leads to visual-spatial image-schemas and these, in turn, provide the basis for language acquisition. Mandler proposed that perceptual analytic processes occurring during infancy (as early as three months) “re-describe” perceptual information into conceptual primitives, which in turn underlie the later acquisition of the relational aspects of language. The following abstract from Mandler (1992b) provides a synopsis of the tenets of her position:

The theory proposes that perceptual analysis re-describes perceptual information into *meanings* [italics added] that form the basis of an accessible conceptual system. These early meanings are represented in the form of image-schemas that abstract certain aspects of the spatial structure of objects and their movements in space. Image-schemas allow infants to form concepts such as animate and inanimate objects, agents, and containers. It is proposed that this form of representation serves a number of functions, including providing a vehicle for simple inferential and analogical thought, enabling the imitation of actions of others, and providing a conceptual basis for the acquisition of the relational aspects of language. (p. 273)

The critical feature of Mandler’s theory is *perceptual analysis*, which more recently (Mandler, 2004) she calls perceptual *meaning* analysis to emphasize that it is a framework of *meanings* that is extracted by the process. Within the theory, perceptual meaning analysis “re-describes” (recodes) perceptual information (both visual and kinesthetic) into spatial meanings, and thus initiates the beginnings of concept formation. Mandler (1992b) further proposed that the “re-description” process begins whenever the infant attentively “notices” (not merely looks at) some aspect of the environmental/bodily stimulus array. She indicates that the re-description of perceptual information results in a simplified form of information that is of less detail, but of “*distilled meaning*” (Mandler, 1992b, p. 277). Mandler did not propose brain mechanisms which might account for the re-description process, or how the distilled meanings come about. It is suggested that both the re-description and distillation of perceptual meaning analysis is precisely what the cerebro-cerebellar decomposition/re-composition of visual-spatial experience and its blending with situation-specific sound patterns accomplish. That is, in agreement with Baddeley, Gathercole, and Papagno (1998), not only did the phonological loop emerge out of this cerebro-cerebellar process in phylogeny, but as Baddeley et al. also concluded, the cerebro-cerebellar process guides the emergence of the phonological loop in ontogeny: “the phonological

loop component of working memory has evolved as a system for supporting language learning” (p. 170).

*A Non-Human Primate Model for the Cerebro-Cerebellar  
Origins of the Phonological Loop*

The foregoing developmental and evolutionary scenario will now be used to describe a detailed non-human primate model of visual-spatial working memory and vocalization as it may have occurred in *Homo habilis* approximately one and half million years ago. While the capuchin monkey is not in direct line with *Homo habilis*, several who have extensively studied the capuchin’s spontaneous tool selection, bipedal transport, and highly refined manipulation of stone tools in nut-cracking have argued that the capuchin is an ideal model for early human evolution (Fragaszy and Cummins–Sebree, 2005; Fragaszy, Visalberghi, and Fedigan, 2004; Visalberghi, Addessi et al. 2009; Visalberghi, Spagnoletti et al. 2009; Westergaard, 1995). Within the overall framework of these arguments Fragaszy and Cummins–Sebree (2005) supplied abundant research evidence that the capuchin provides a good model for the study of the spatial reasoning across eras of human prehistory. This places the visual-spatial working memory capacities of the capuchin monkey somewhere near the very beginnings of human language evolution.

Moreover, Obayashi et al. (2007), in studying whether monkeys can master remote visual-spatial operations under different sets of rules, found that monkeys are able to learn to organize abstract sequential operations under context-related rules and be able to understand causal relationships, thus implying the use of abstract cognitive representations (mental models). And, since Obayashi et al. found that cerebro-cerebellar loops mediate the learning of these abstract sequential operations in monkeys, their visual-spatial processes appear to employ cerebellar internal model blending to accomplish new and challenging tasks as found by Imamizu, Higuchi, Toda, and Kawato (2007). Therefore Obayashi et al’s findings, along with the several above findings of lab and field studies of capuchins, are particularly pertinent to the use of monkeys as a rough model of *Homo habilis*, a model that sheds direct light, perhaps, on the earliest beginnings of the recent three- to four-fold evolutionary expansion of the cerebellum toward human language proposed by Leiner, Leiner, and Dow (1986).

*The Adaptive Rise to the Phonological Loop*

Figure 2 is a generalized representation of the capuchin nut-cracking sequence observed in wide variety of analytic field studies (Fragaszy, Greenberg et al. 2010; Fragaszy, Izar, Visalberghi, Ottoni, and Gomes De Oliveira, 2004; Fragaszy, Pickering

et al. 2010). The sequence in Figure 2 begins the nut-cracking sequence after the capuchin has already (1) searched for and (2) carefully selected appropriate stone “hammers” which will crack the nuts, and (3) bipedally transported the hammer(s) to a (4) selected “anvil” area where the nuts will be cradled during cracking efforts. Since the capuchin is serving as a theoretical model for *Homo habilis*, the sequence will hereafter be referred to as that of *capuchin–Homo habilis* to metaphorically emphasize the suggested parallels between their two evolutionary sequences.

The actual nut-cracking sequence begins at the left in Figure 2 with an *internal representation* (I) of the prospective goal (obtaining kernels from a palm nut) in the capuchin’s visual-spatial working memory. The precise nature of this internal representation is an important issue. Here, the notion of the internal representation as the starting point in the sequence is based upon Baddeley and Andrade’s (2000) proposal that the mental modeling imagery of goals in the visual-spatial working memory of early humans was selectively adaptive because it helped in the prediction of future events and therefore in the planning of action. This idea is in complete agreement with Fuster’s (2008) definition of working memory as, “the temporary retention of information — sensory or other — for the performance of a prospective act to solve a problem or to attain a goal” (p. 138). Once learned through the repetitive experience of meeting the novel challenge of cracking each new nut, the internal representation contains the entire sequence of associations necessary to initiate the working memory imagery and actions

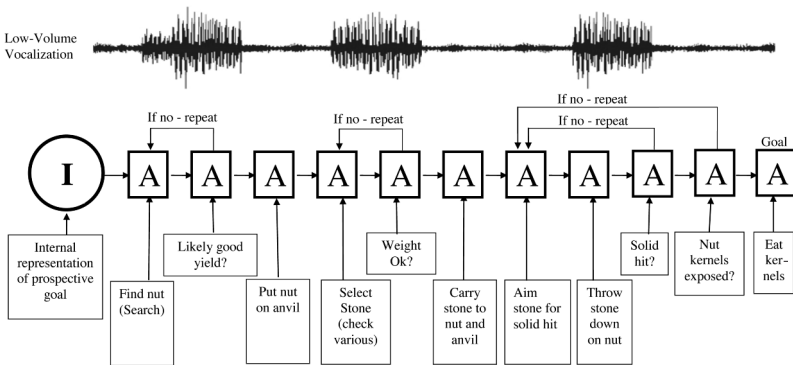


Figure 2: Generalized sequence of actions driven by visual-spatial working memory during nut-cracking in capuchins. Following the internal representation (I) of a prospective, mentally modeled goal in visual-spatial working memory, the orderly series of actions (A) is mediated by cerebro-cerebellar loops. Across the top is low-volume vocalization which is hypothesized to be associated with capuchin sub-vocalization while visual-spatial working memory and manipulation is working through the ever-new challenges of the “If no — repeat” sub-routines of nut-cracking (see footnote 3).



necessary to achieving the prospective goal (see Fuster, 2008, p. 364, Figure 8.5 for a detailed discussion of the organization of such a sequence). The learning of the entire internal representation takes about two years in capuchins and is mediated by changes in cerebro-cerebellar re-organization or blending occurring over thousands of learning attempts (repetitions). Why is cerebro-cerebellar re-organization or blending involved in the establishment of the internal representation? Fuster (2008) emphasized the role of novelty or the confronting of new, challenging circumstances (as did Imamizu, Higuchi, Toda, and Kawato, 2007) in the adaptive evolution of working memory in the prefrontal cortex. Fuster pointed out that in working memory delay tasks using monkeys “each trial must be treated by the animal as unique and independent from previous trials. The game is old, but each play is new” (2008, p. 366). Likewise, as the capuchin goes through each “trial” in the sequence in Figure 2, the nut, its position on the anvil, and the required aim and force of the stone hammer throw, constitute a randomly new event (Liu et al., 2011).

Proceeding on from the internal representation in Figure 2, the typical capuchin sequence of nut-cracking *actions* (A) unfolds to the right. Learned imagery sequences in working memory guide both the linear series of actions and those of the four sub-routines, the latter each indicated by a question and an “If no — repeat” internally mediated instruction. Thus, each time a palm nut is satisfactorily cracked, reaching the goal requires *the repetition of an orderly sequence of visual-spatial imagery involving orderly steps of decision-making*. This orderly sequence constitutes a *syntax of actions* (Fuster, 2008; Lashley, 1951), but, since the path of the sequence of actions is guided by a sequence of imagery in working memory, it is more importantly an ordering or syntax in cognitive decision-making. The cerebro-cerebellar formation of this syntax requires thousands of repetitions through the entire sequence of actions and thus comes to constitute the internal representation (I) of the sequence. This syntax of cognitive/visual-spatial decision making in capuchins involves precisely the same process as the “abstract sequential operations” that Obayashi et al. (2007) found in monkeys and which they referred to as “a relatively sophisticated system of internal representation [mental models] in the absence of language” (abstract, p. 389).

*Vocalizations Paralleling the Nut-Cracking Sequence: The Earliest Sub-Vocal Moments of the Phonological Loop*

Dorothy Fragaszy, whose extensive research on capuchins is cited above, has indicated that capuchins in general “comment” on their manipulative activities with low-volume vocalizations (personal communication, 2011). Running across the top of the action sequence in Figure 2 is a line representing those ongoing low-volume vocalizations.

Why would capuchins pair low-volume vocalizations with tool manipulation? A definitive answer to this question is not yet known. However, in the meantime, the following hypothesis is proposed. Desmond, Gabrieli, Wagner, Ginier, and Glover (1997) and Marvel and Desmond (2010) concluded that *sub-vocal speech enhances the effectiveness of working memory in complex tasks*, and that sub-vocal speech accomplishes this by operating through cerebro-cerebellar loops which correct errors in vocalization and tool-use computations emanating from Broca's area. It is proposed that the low-volume vocalizations heard in capuchins during manipulative activities are an indication of the presence of ongoing *sub-vocalizations* which, as Desmond et al. and Marvel and Desmond found in humans, served that same purpose in capuchins, namely, *enhanced effectiveness of working memory in complex tasks*. Thus, in capuchins, the low-volume vocalizations accompanying tool manipulation may aid in negotiating the long delays and sub-routines (especially during early learning) in the nut-cracking sequence depicted in Figure 2 and thereby provide a significant adaptive advantage. It is further hypothesized, therefore, that in capuchin–Homo habilis, cerebellar internal models of orderly sequences of visual-spatial working memory and actions associated with sequences of tool use depicted in Figure 2 were *blended* with cerebellar internal models of sub-vocalizations via the same processes proposed by Imamizu, Higuchi, Toda, and Kawato (2007) and described above.

It is proposed that this blending resulted in adaptive vocal coding (vocal tagging and filing) of socially common actions and cognitive syntaxes in the long-term memory of capuchin–Homo habilis and thus began the gradual emergence of the phonological staging and manipulation of imagery sequences in active (working) memory as suggested above by Baddeley, Gathercole, and Papagno (1998). Baddeley (1992) suggested a similar evolutionary scenario for working memory:

Working memory stands at the crossroads between memory, attention, and perception. In the case of the slave systems, the phonological loop, for example, probably represents an evolution of the basic speech perception and production systems to the point at which they can be used for active memory. (p. 559)

Directly supporting Baddeley's scenario, it is argued that the vocal tagging and filing in long-term memory during the evolutionary emergence of the phonological loop was the result of the following two interrelated contributions of cerebro-cerebellar collaboration. First, upon encountering new, challenging environmental demands which pressed the limits of then-existing stone tool technology, cerebellar internal models gradually decomposed/re-composed visual-spatial experience associated with situation-specific actions ([A] in Figure 2), and parallel situation-specific vocalizations into further decompositions/re-organizations of cerebellar internal models (Flanagan et al. 1999; Haruno, Wolpert, and Kawato, 2003; Nakano et al. 2002) which, when blended, selected toward

new, uniquely human syntactic orders of language features. Second, these new vocal differentiations in evolving visual-spatial working memory served as an increasingly larger system of associative internal and social vocal tags (Fuster, 2008, pp. 249–251). Cerebellar internal models associated with this proliferation of vocal tags gradually selected toward a rehearsal loop in working memory (phonological loop) enabling capuchin–*Homo habilis* to negotiate longer and longer delays in task sequences represented in working memory (Fuster, 2008, p. 185). It is suggested that the cerebellar internal models representing these new differentiations in vocal tagging may have first appeared as “silent” vocalizations (sub-vocalizations) which allowed capuchin–*Homo habilis*, as Marvel and Desmond (2010) suggested in their research on humans, “to represent, maintain, and organize task-relevant information and conscious thoughts whenever talking through a solution to a problem or complex sequence of tool construction or use is beneficial” (p. 273). The ability to negotiate longer task delays in this manner appears to have been the driving force that eventually led *Homo habilis* to develop more articulated multiple-component types of tools (Ambrose, 2001), which, by expanding the sequence and sub-routines of the type in Figure 2, then in turn further articulated vocal tagging through the blending process described by Imamizu et al. (2007). According to this view, new stone-tool technology was first imagined in visual-spatial imagery within longer task delays and with the aid of sub-vocalization. Since, as stated above, the cerebro-cerebellar formation of longer task delays (and therefore the capacity to imagine new stone-tool technology as an internal representation, [I] in Figure 2) required long periods of repetitions through the sequence of actions [A], it is argued that emerging internal representations acquired through repetitions in working memory led advances in stone-tool technology; stone-tool technology per se did not lead advances in working memory.

It is suggested that the gradual, generation-by-generation cerebro-cerebellar blending of visual-spatial behavioral sequences paired with the personal/social access system of parallel vocalization sequences (the phonological loop) would have provided not only a way of allowing *Homo habilis* to manipulate detailed cause-and-effect relationships in personal visual-spatial experience but also a way of communicating that visual-spatial cause-and-effect imagery to others in some detail (see footnote 3). The gradual emergence of an infinitely partitionable working memory and, at the same time, a *socially sharable* working memory about detailed cause-and-effect relationships would have been of enormous selective advantage. This would have been the working memory first envisioned by Tooby and DeVore (1987) in their concept of the uniquely human advantage, the “cognitive niche,” wherein early humans became cognitively superior to all prey and predator species: “We accomplish[ed] this by conceptually abstracting from a situation a model of what manipulations are necessary to achieve proximate goals that correlate with fitness. These highly orchestrated

and intricate situation-specific sequences of behavior are cognitively organized” (p. 209). Pinker (2010) supportively elaborated on Tooby and DeVore’s notion of the evolution of the cognitive niche of early humans in some detail. It now appears that the evolutionary expansion of cerebro-cerebellar blending of visual-spatial working memory with vocalization was the driving mechanism behind the emergence of the cognitive niche in early humans.

### *Conclusion*

Imamizu et al’s (2000, 2003, 2007; Imamizu and Kawato, 2009) imaging studies on the cerebro-cerebellar blending of internal models can serve as a model for the mechanisms behind the last million or so years of the evolution of language and other abstract reasoning processes. These cerebro-cerebellar mechanisms, when combined with the work of Baddeley and his colleagues, move the study of language away from a strictly internal analysis of the structure and meaning of language to one that places language evolution within the larger intertwined evolution of the components of working memory. It is argued that the last million years of the evolution of cerebro-cerebellar blending of visual-spatial working memory with vocalization in *Homo habilis* (or perhaps even earlier) was the driving mechanism behind the evolution of the human “cognitive niche” first proposed by Tooby and DeVore (1987).

Since the brain mechanisms underlying the proposals in this article are based upon Imamizu, Higuchi, Toda, and Kawato’s (2007) experimental imaging procedures, it is suggested that these procedures be adapted to imaging studies of cerebro-cerebellar involvement in acquisition of new vocabulary as studied by Baddeley, Gathercole, and Papagno (1998). Marvel and Desmond (2010, in press) have performed imaging studies that are closely related to this suggestion, and which could be adapted directly to Baddeley et al.-type studies within the framework of Imamizu et al.’s procedures. Further adaptations of Marvel and Desmond’s (2010, in press) imaging studies would be to apply them to the study of the low-volume vocalizations of capuchins during nut-cracking. It is hypothesized that capuchin vocalizations during tool manipulation are indications of sub-vocalization, and that sub-vocalization activity would be most intense and most varied during the four “If no — repeat” sub-routines illustrated in Figure 2.

A strong illustration of a gateway for the working memory-based study of language evolution was the work of the late Charles Hockett. Just over 50 years ago in his now classic article that appeared in *Scientific American*, Hockett (1960) proposed that language evolved from the *blending* of fractionated call systems of pre-humans. He was convinced that the selective blending of combinations of different calls in the pressure of new, challenging environmental circumstances led to language evolution. In Hockett’s time, he was unable to point to

brain systems or brain mechanisms to support his contentions about some sort of brain-based process of blending that would have bestowed advantage. Detailed theoretical models and laboratory studies of working memory and of the cognitive and language functions of cerebro-cerebellar collaboration were off in the future. But the theories and findings of these new sciences have indicated that Hockett was not far off the mark in his conjectures (see footnote 2).

## References

- Aboitiz, F., Garcia, R., Bosman, C., and Brunetti, E. (2006). Cortical memory mechanisms and language origins. *Brain and Language*, 98, 40–46.
- Ambrose, S. (2001, March 2). Paleolithic technology and human evolution. *Science*, 291, 1748–1753.
- Baddeley, A. (1992, January 31). Working memory. *Science*, 255, 556–559.
- Baddeley, A., and Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General*, 129, 126–145.
- Baddeley, A., Gathercole, S., and Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105, 158–173.
- Desmond, J., Gabrieli, J., Wagner, A., Ginier, B., and Glover, G. (1997). Lobular patterns of cerebellar activation in verbal working-memory and finger-tapping tasks as revealed by functional MRI. *Journal of Neuroscience*, 17, 9675–9685.
- Flanagan, R., Nakano, E., Imamizu, H., Osu, R., Yoshioka, T., and Kawato, M. (1999). Composition and decomposition of internal models in learning under altered kinematic and dynamic environments. *Journal of Neuroscience*, 19, 1–5.
- Fragaszy, D., and Cummins-Sebree, S. (2005). Relational spatial reasoning by a nonhuman: The example of capuchin monkeys. *Behavioral and Cognitive Neuroscience Reviews*, 4, 282–306.
- Fragaszy, D., Greenberg, R., Visalberghi, E., Ottoni, E., Izar, P., and Liu, Q. (2010). How wild bearded capuchin monkeys select stones and nuts to minimize the number of strikes per nut cracked. *Animal Behavior*, 79, 1–10.
- Fragaszy, D., Izar, P., Visalberghi, E., Ottoni, E., and Gomes De Oliveira, M. (2004). Wild capuchin monkeys (*Cebus libidinosus*) use anvils and stone pounding tools. *American Journal of Primatology*, 64, 359–366.
- Fragaszy, D., Pickering, T., Liu, Q., Izar, P., Ottoni, E., and Visalberghi, E. (2010). Bearded capuchin monkeys' and a human's efficiency at cracking palm nuts with stone tools: Field experiments. *Animal Behavior*, 79, 321–332.
- Fragaszy, D., Visalberghi, E., and Fedigan, L. (2004). *The complete capuchin*. Cambridge: Cambridge University Press.
- Fuster, J. (2008). *The prefrontal cortex* (fourth edition). London: Academic Press.
- Haruno, M., Wolpert, D., and Kawato, M. (2003). Hierarchical MOSAIC for movement generation. In T. Ono, G. Matsumoto, R.R. Llinas, A. Bethoz, R. Norgren, H. Nishijo, and R. Tamura (Eds.), *Excepta Medica International Congress Series* (Volume 1250, pp. 575–590). Amsterdam: Elsevier Science.
- Hockett, D. (1960). The origin of speech. *Scientific American*, 203(3), 88–96.
- Imamizu, H., Higuchi, S., Toda, A., and Kawato, M. (2007). Reorganization of brain activity for multiple internal models after short but intensive training. *Cortex*, 43, 338–349.
- Imamizu H., and Kawato, M. (2009). Brain mechanisms for predictive control by switching internal models: Implications for higher-order cognitive functions. *Psychological Research*, 73(4), 527–544.
- Imamizu, H., and Kawato, M. (in press). Cerebellar internal models: Implications for dexterous use of tools. *Cerebellum*.
- Imamizu, H., Kuroda, T., Miyachi, S., Yoshioka, T., and Kawato, M. (2003). Modular organization of internal models of tools in the human cerebellum. *Proceedings of the National Academy of Science*, 100(9), 5461–5466.

- Imamizu, H., Miyauchi, S., Tamada, T., Sasaki, Y., Takino, R., Pütz, B., Yoshioka, T., and Kawato, M. (2000). Human cerebellar activity reflecting an acquired internal model of a new tool. *Nature*, 403, 192–195.
- Lashley, K. (1951). The problem of serial order in behavior. In L.A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–131). New York: Wiley.
- Leiner, H., Leiner, A., and Dow, R. (1986). Does the cerebellum contribute to mental skills? *Behavioral Neuroscience*, 100, 443–454.
- Leiner, H., Leiner, A., and Dow, R. (1989). Reappraising the cerebellum: What does the hindbrain contribute to the forebrain? *Behavioral Neuroscience*, 103, 998–1008.
- Liu, Q., Frigaszy, D., Wright, B., Wright, K., Izar, P., and Visalberghi, E. (2011). Wild bearded capuchin monkeys (*Cebus libidinosus*) place nuts in anvils selectively. *Animal Behavior*, 81, 297–305.
- Mandler, J. (1992a). How to build a baby II: Conceptual primitives. *Psychological Review*, 99, 587–604.
- Mandler, J. (1992b). The foundations of conceptual thought in infancy. *Cognitive Development*, 7, 273–285.
- Mandler, J. (2004). *The foundations of mind: Origins of conceptual thought*. Oxford: Oxford University Press.
- Mandler, J. (in press). On the spatial foundations of the conceptual system and its enrichment. *Cognitive Science*.
- Marvel, C., and Desmond, J. (2010). Functional topography of the cerebellum in verbal working memory. *Neuropsychology Review*, 20, 271–279.
- Marvel, C., and Desmond, J. (in press). From storage to manipulation: How the correlates of verbal working memory vary demands on inner speech. *Brain & Language*.
- Nakano, E., Flanagan, J., Imamizu, H., Rieko, O., Yoshioka, T., and Kawato, M. (2002). Composition and decomposition learning of reaching movements under altered environments: An examination of the multiplicity of internal models. *Systems and Computer in Japan*, 33(11), 80–94.
- Obayashi, S., Matsumoto, R., Suhara, T., Nagai, Y., Iriki, A., and Maeda, J. (2007). Functional organization of the monkey brain for abstract operation. *Cortex*, 43, 389–396.
- Obayashi, S., Suhara, T., Nagai, Y., Maeda, J., Hihara, S., and Iriki, A. (2002). Macaque prefrontal activity associated with extensive tool use. *Neuroreport*, 13, 2349–2354.
- Passingham, R. (1975). Changes in the size and organization of the brain in man and his ancestors. *Brain, Behavior and Evolution*, 11, 73–90.
- Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. *PNAS*, 107(2), 8993–8999.
- Stephan, H., and Andy, O. (1969). Quantitative comparative neuroanatomy of primates. In J.M. Petras and C.R. Noback (Eds.), *Comparative and evolutionary aspects of the vertebrate nervous system* (Annals, Volume 167, pp. 370–387). New York: New York Academy of Sciences.
- Tooby, J., and DeVore, I. (1987). The reconstruction of hominid behavioral evolution through strategic modeling. In W.G. Kinzey (Ed.), *The evolution of human behavior: Primate models* (pp. 183–237). Albany: State University of New York Press.
- Vandervert, L. (1997). The evolution of Mandler's conceptual primitives (images schemas) as neural mechanisms for space–time simulation structures. *New Ideas in Psychology*, 15, 105–123.
- Vandervert, L. (2003). How working memory and cognitive modeling functions of the cerebellum contribute to discoveries in mathematics. *New Ideas in Psychology*, 21, 159–175.
- Vandervert, L. (2009). The emergence of the child prodigy 10,000 years ago: An evolutionary and developmental explanation. *Journal of Mind and Behavior*, 30, 15–32.
- Vandervert, L. (in press). How the blending of cerebellar internal models can explain the evolution of thought and language. *Cerebellum*.
- Vandervert, L., Schimpf, P., and Liu, H. (2007). How working memory and the cognitive functions of the cerebellum collaborate to produce creativity and innovation. *Creativity Research Journal*, 19, 1–18.
- Visalberghi, E., Addessi, E., Truppa, V., Spagnoletti, N., Ottoni, E., Izar, P., and Frigaszy, D. (2009). Selection of effective stone tools by wild capuchin monkeys. *Current Biology*, 19, 213–217.
- Visalberghi, E., Spagnoletti, N., Ramos da Silva, E., Andrade, F., Ottoni, E., Izar, P., and Frigaszy, D. (2009). Distribution of potential suitable hammers and transport of hammer tools and nuts by wild capuchin monkeys. *Primates*, 50, 95–104.

- Washburn, S., and Harding, R. (1970). Evolution of primate behavior. In F.O. Schmitt (Ed.), *The neurosciences: Second study program* (pp. 39–47). New York: Rockefeller University Press.
- Westergaard, G. (1995). The stone-tool technology of capuchin monkeys: Possible implications for the evolution of symbolic communication in hominids. *Symbolic Aspects of Early Technologies*, 27, 1–9.

