Creativity Research Journal
2007, Vol. 19, No. 1, 1–18

FEATURED ARTICLE

How Working Memory and the Cerebellum Collaborate to Produce Creativity and Innovation

Larry R. Vandervert
American Nonlinear Systems

Paul H. Schimpf and Hesheng Liu
Washington State University–Spokane

ABSTRACT: It is proposed that (a) creativity and innovation are the result of continuously repetitive processes of working memory that are learned as cognitive control models in the cerebellum, (b) that these cerebellar control models consist of multiple-paired predictor (forward) models within the MOdular Selection and Identification for Control (MOSAIC) and hierarchical MOSAIC (HMOSAIC) cerebellar architectures that explore and test problem-solving requirements, and (c) when resulting newly formed predictor/controller models are fed forward to more efficiently control the operations of working memory, they lead to creative and innovative problem solving (including the experiences of "insight" and "intuition"). Within this framework, three of Einstein's classic autobiographical accounts of creative discovery are analyzed. It is concluded that the working memory/cerebellar explanation of creativity and innovation can begin to tie together: (1) behavioral and neuroimaging studies of working memory, (2) behavioral, clinical and neuroimaging studies of the cognitive functions of the cerebellum, and (3) autobiographical accounts of creativity. It is suggested that newly developed electromagnetic inverse techniques will be a necessary complement to functional brain imaging studies to further establish the validity of the theory.

Elsewhere, Vandervert (2003a, 2003b) sketched a preliminary theory of how mathematical discovery, creativity and innovation arise through the collaboration of working memory and the cognitive functions of the cerebellum. Working memory may be thought of as the "online" cognitive consciousness through which one acquires new knowledge, solves problems and formulates and acts on current goals (e.g., Baddeley, 1992; Baddeley & Logie, 1999; Baddeley & Andrade, 1998; Cowan, 1999; Ericsson & Kintsch, 1995). The cognitive functions of the cerebellum consist of the modeling of repetitive cognitive functions of the cerebral cortex (including those of working memory) in ways that, when fed back to the cerebral cortex, increase the speed, efficiency, and adaptability of the original cerebral functions (e.g., Desmond & Fiez, 1998; Haruno, Wolpert, & Kawato, 1999, 2001; Houk & Wise, 1995; Imamizu et al., 2003; Ito, 1997; Leiner & Leiner, 1997; Leiner et al., 1986, 1989; Wolpert, Doya, & Kawato, 2003). Working memory and the cognitive functions of the cerebellum will be described more fully below.

Purpose

In the present article, further support for the working memory/cerebellar theory of creativity and innovation is provided through the following: (a) updated and ex-

Address correspondence to Larry R. Vandervert, American Nonlinear Systems, 1529 W. Courtland, Spokane, WA 99205. E-mail: LVandervert@aol.com

panded neuroimaging research on working memory (e.g., Chein, Ravizza, & Fiez, 2003), and the cognitive functions of the cerebellum (Oztop, Wolpert, & Kawato, 2005; Wolpert et al., 2003), (b) the perceptual analytic origin of image-schemas (conceptual primitives; Mandler, 2004), and, on the basis of a and b, (c) more thoroughly detailed analyses of Einstein's three classic autobiographical accounts of creative and innovative thinking.

Three interrelated arguments are presented in this article. First, it is argued that in the same way that cerebellar models for the control of repetitive bodily movements are learned and fed back to the cerebral cortex, models for the repetitive cognitive processes of working memory are learned in the cerebellum and fed back to working memory, making its attentional, visuospatial, and language functions significantly faster, more efficient and adaptive (e.g., Akshoomoff, Courchesne, & Townsend 1997; Cabeza & Nyberg, 2000; Chein et al., 2003; Desmond & Fiez, 1998; Doya, 1999; Imamizu, Kuroda, Miyauchi, Yoshioka, & Kawato, 2003; Ito, 1993, 1997; Leiner et al., 1986, 1989). Second, because the components of working memory contain the attributes of conscious awareness and imagery (Baddeley, 1993; Baddeley & Andrade, 1998; Baddeley & Hitch, 1974; Cowan, 1999; Ericsson & Kintsch, 1995; Ericsson & Delaney, 1999; Teasdale et al., 1995), it is argued that, although the construction of cerebellar models itself is not accessible to consciousness, the fed forward, adaptive cerebellar modifications of working memory provide for consciously accessible products of creativity and innovation. Finally, it is a contention of this article that the unique experiences of creative discovery and innovation (including the experience of "insight," and what has traditionally been called "intuition" in the arts and sciences) are the result of newly formed forward (predictor) models within the MOdular Selection and Identification for Control (MOSAIC) and hierarchical MOSAIC (HMOSAIC) cerebellar architectures (Haruno et al., 2001; Wolpert et al., 2003) when they make new connections among layers of the HMOSAIC architecture. We now turn to a description of working memory.

Working Memory: The Ongoing Stream of Cognitive Consciousness

Working memory consists of a collection of cognitive functions that is engaged whenever we are doing what most people would call "thinking"—that is, whenever we are carrying out both simple and complex everyday cognitive tasks. When, for example, we read an article in the newspaper (or a scientific journal), mentally rearrange the furniture in our living room to make room for a new sofa, compare and contrast the attributes of several new cars before making a purchase, give directions to our home, or even make change at the grocery store, we are using working memory (Miyake & Shah, 1999, chap. 1). Another interesting, but not often cited, example of working memory occurs whenever a person responds to items in a psychological test of, for example, personality, intelligence, or, of course, creativity. Working memory is also at "work" in the high-level performances of experts in all fields (mathematical calculators, chess and music masters, on-the-feet thinking of seasoned university professors, and so on; Ericsson & Kintsch, 1995; Ericsson, 2003a). Because working memory in such experts involves a good deal of readily accessible long-term memory material that has been acquired over many years of study and practice, Ericsson and Kintsch (1995) refer to their approach to working memory as *long-term working memory*.[1]

Cowan (1999) provided the following definition of working memory; it is in close general agreement with definitions provided by other working-memory theorists (see Miyake & Shah, 1999, pp. 450–452):[2]

[1]It is important to note that Ericsson and his colleagues found that experts and exceptional performers deliberately avoid *automaticity*, or preprogrammed skill patterns in their abilities by engaging in problem solving about how to constantly improve their skills at each level of mastery (Ericsson, 2002, 2003a). Quite to the contrary to automaticity, Ericsson (2002, 2003a) has shown how *deliberate practice* (defined as practice aimed toward constantly elevated levels of performance) leads to the development of high-level skills in both responding to novel situations and manufacturing novel behavior. This deliberate practice of constantly new levels of problem-solving behavior, like all other regularly practiced routines, is itself apparently made constantly more efficient through the collaboration of working memory and the cerebellum. Such developments of increasing levels of problem solving are supported by the hierarchical levels of the HMOSAIC cerebellar architecture.

[2]Miyake and Shah (1999) combined ten definitions of working memory provided by leading theorists in the following all-encompassing definition:

> Working memory is those mechanisms or processes that are involved in the control, regulation, and active maintenance of

> Working memory refers to cognitive processes that retain information in an unusually accessible state, suitable for carrying out any task with a mental component. The task may be language comprehension or production, problem solving, decision making, or other thought. (Cowan, 1999, p. 62)

The key to understanding the cognitive processes of working memory is that they retain information from memory stores (both short-term and long-term memory) within a mentally graspable range during thought. Nearly all working-memory theorists agree that the working-memory components that accomplish this maintenance task include a central executive function, and two slave functions: a visuospatial sketchpad and a speech loop (e.g., Baddeley, 1992; Baddeley & Logie, 1999; Cowan, 1999).

As an illustration of the functions of the components of working memory any of the everyday examples cited earlier could be used. We use the example of reading a newspaper to take a look at the operation of working memory. This example will help lay the groundwork for the reader for our later analysis of Einstein's autobiographical accounts. First, *attentional control* in reading and thinking about various newspaper articles is carried on by working memory's central executive functions. Attentional functions of the central executive supervise, schedule and integrate information from different sources. The visuospatial sketchpad and the speech loop are, respectively, manipulation and rehearsal processes for retaining appropriate visuospatial images and speech information that are needed for the on-line comprehension, decision making, and thinking about the contents of the various newspaper articles. To maintain information in a conscious, "on-line" state so that these mental tasks can be completed, the central executive employs the visuospatial sketchpad and the speech loop in a continual process of *repetitive manipulation, rehearsal and updating*.

Before going on, it will be helpful to again make note that the leading argument of this article is that, like the repetitive components of bodily movements, it is the above *repetitive* actions (manipulation and rehearsal) and interactions of the components of working memory that are modeled in the cerebellum and subsequently fed back to working memory making its operations faster, more efficient, and more adaptive (e.g., Ito, 1997; Leiner et al., 1986, 1989). Neuroimaging studies have confirmed that these working memory processes can be associated with various areas of both the cerebral cortex and the cerebellum (e.g., Cabeza & Nyberg, 2000; Chein et al., 2003; Desmond & Fiez, 1998; Fiez et al., 1996). Therefore, there is little doubt that whatever working memory accomplishes, it does it through collaboration with the cerebellum.

In summary, working memory is not just about "memory." Working memory is where thinking, problem solving, daydreaming, expert and exceptional performance take place (see Footnote 2). Goldman-Rakic (1992) said the following of working memory:

> The combination of moment-to-moment awareness and instant retrieval of archived information constitutes what is called working memory, perhaps the most significant achievement of human evolution. It enables humans to plan for the future and to string together thoughts and ideas, which has prompted Marcel Just and Patricia Carpenter of Carnegie Mellon University to refer to working memory as 'the blackboard of the mind.' (p. 111)

In this overall vein of advanced brain processes, we believe working memory is also where creativity and innovation are born. But the full story of working memory that makes it the most significant achievement of human evolution and explains creativity and innovation involves more than its traditional components. We believe creativity and innovation in working memory also necessarily involve contributions from the cerebellum.

The New Perception of the Cerebellum: The Cognitive Functions of the Cerebellum

In the course of everyday repetitive mental/physical activities (e.g., driving a car, playing computer games, or playing basketball), a person becomes able

---

task-relevant information in the service of complex cognition, including novel as well as familiar, skilled tasks. It consists of a set of processes and mechanisms and is not a fixed "place" or "box" in the cognitive architecture. It is not a completely unitary system in the sense that it involves multiple representational codes and/or different subsystems. Its capacity limits reflect multiple factors and may even be an emergent property of multiple processes and mechanisms involved. Working memory is closely linked to [long-term memory] LTM, and its contents consist primarily of currently activated LTM representations, but can also extend to LTM memory representations that are closely linked to activated retrieval cues and, hence, can be quickly activated. (p. 450)

to execute the required tasks more quickly and precisely and in novel ways. The development of fast, highly controlled, problem-solving expertise in all mental/physical skills (sports, recreation, occupational, military) relies on this fact. It is well established that, across the board, these increases in efficiency and adaptability are the result of control routines that are learned in the cerebellum and subsequently fed back to control improved timing and sequencing of the operations of the movement-generating (motor) portions (and other related parts) of the brain's cerebral cortex (e.g., Bloedel, Dichgans, & Precht, 1985; Ito, 1984a, 1997; Kornhuber, 1974; Thach, 1996). These increases in efficiency and adaptability do, in part, lead to automaticity of behavior (e.g., Kihlstrom, 1987). However, they equally lead to the development of creative and innovative cerebellar control routines for the cerebral cortex. More will be said below in the section on *Cerebellar Role in Manipulation of Thought: Conscious and Unconscious Control in Working Memory* of how these two seemingly contrary control outcomes (automaticity and novelty) complement one another.

In the last 20 years, understandings of the cerebellum have moved far beyond the earlier, more traditional idea that its functions are limited to motor control. A number of newer and converging lines of research and theory, especially those arising from neuroimaging studies, demonstrate that the cerebellum provides a fast computational system for the timing, sequencing and modeling aimed at the rapid manipulation of both motor and cognitive processes, including working memory (e.g., Akshoomoff et al., 1997; Cabeza & Nyberg, 2000; Chein et al., 2003; Desmond & Fiez, 1998; Doya, 1999; Houk & Wise, 1995; Haruno et al., 1999; Haruno, Wolpert, & Kawato, 2001; Imamizu et al., 2003; Ito, 1993, 1997; Ivry, 1997; Leiner & Leiner, 1997; Leiner et al., 1986, 1989; Schmahmann, 1997).[3]

Early Foundational Arguments Concerning Cognitive Functions of the Cerebellum

It appears that human memory and working memory assumed their present systems of components (including the central executive, visuospatial sketchpad, speech loop discussed earlier) during the last million years of hominid evolution (e.g., Baddeley, 1993; Baddeley & Andrade, 2000; Klein, Cosmides, Tooby, & Chance, 2002). A decade and a half before Klein et al. (2002) proposed their somewhat detailed explanation of the selective evolution of memory, Leiner et al. (1986, 1989) speculated that the cerebellum contributed to such mental skills. In their foundational articles on the cognitive functions of the cerebellum, Leiner et al. pointed out that during the same million years cited by Klein et al., the cerebellum enlarged by an astonishing three to four times, and that the cerebro-cerebellar system (the extensive hardwiring between the cerebral cortex and the cerebellum) became more elaborately and extensively interconnected. Central to the arguments of the present article, they also proposed that the cerebellum had, during this evolutionary time, become involved in the manipulation of ideas:

> It has often been remarked that an explanation is required for the threefold to fourfold increase in the size of the cerebellum that occurred in the last million years of evolution (Washburn & Harding, 1970). If the selection pressure has been strong for more cerebellum in the human brain as well as for more cerebral cortex, the interaction between the cerebellum and the cerebral cortex should provide some important advantages to humans… a detailed examination of cerebellar circuitry suggests that its phylogenetically newest parts may serve as a fast information-processing adjunct of the association cortex and could assist this cortex in the performance of

[3]According to the *computational* scheme for mental models set forth originally by Craik (1943) and extensively elaborated by Johnson-Laird (1983), thought processes construct predictive models that are imitative, small-scale computational representations of the external world that retain the external world's relation-structure. Mental models and cerebellar models we discuss throughout this article may be thought of as the Johnson-Laird type. Craik described how the preserved relation-structure of the model is computationally parallel in its predictive ability to that which it imitates:

> A calculating machine, an anti-aircraft 'predictor', and Kelvin's tidal predictor all show the same ability. In all of these cases, the physical process, which it is desired to predict, is *imitated* [the relation-structure is preserved] by some mechanical device or model which is cheaper, quicker, or more convenient in operation. (1943, p. 52)

In the case of modeling by the cerebellum, models (forward and inverse) are composed from of mental models operating in the cerebral cortex, including those of working memory. Thus, Ito (1997) suggested the relation-structure of the cerebellar model captures the fundamentals of predictive operation of mental models used in the cerebral cortex; the cerebellar model thus makes the operations of the cortical models faster, "neurologically cheaper," and adaptive.

> a variety of manipulative skills, including the skill that is characteristic of anthropod apes and humans, the skillful manipulation of ideas. (Leiner et al., 1986, p. 444)

The skillful manipulation of ideas, of course, is precisely the job of working memory. Then in a follow-up article 3 years later, Leiner et al. (1989) extended their findings and arguments to include the skillful manipulation of language functions:

> We conclude that the phylogenetically newest circuitry in the human cerebro-cerebellar system enables the cerebellum to improve the speed and skill of cognitive and language performance [particularly circuitry connected with Brodmann areas 44 and 45, which constitute part of Broca's language area], in much the same way that the phylogenetically older circuitry enables the cerebellum to improve the speed and skill of motor performance. (p. 999)

See Figure 1.

Leiner et al. (1986, 1989) also proposed that decisional and search skills (functions of the central executive and the same memory control skills proposed by Klein et al., 2002) are learned in the cerebellum through its extensive feedback loop connections with the frontal areas of the brain. The ensuing 15 years of human neuroimaging studies of the activities of the cerebro-cerebellar system has born out Leiner et al.'s (1989) early, preliminary arguments and evidence of extensive reciprocal feedback modulation between the cerebellum on the one hand, and language and a variety of other mental skill areas of the cerebral cortex, including those associated with working memory on the other (Cabeza & Nyberg, 2000; Chein et al., 2003; Desmond & Fiez, 1998; Doya, 1999). Thus, today, there is ample evidence that, indeed, working memory and the cerebellum have, over the course of human evolution, come to collaborate in making all online thought processes increasingly more efficient and adaptive (see Ito, 1997, for a thoroughgoing treatment of the adaptive role of the cerebellum in all levels of human and animal brain function).

In order to understand the logic of how the cerebellum's role in the manipulation of ideas should not be differentiated from its manipulation of movement, Ito (1993, 1997) pointed out that, at the neurological level,

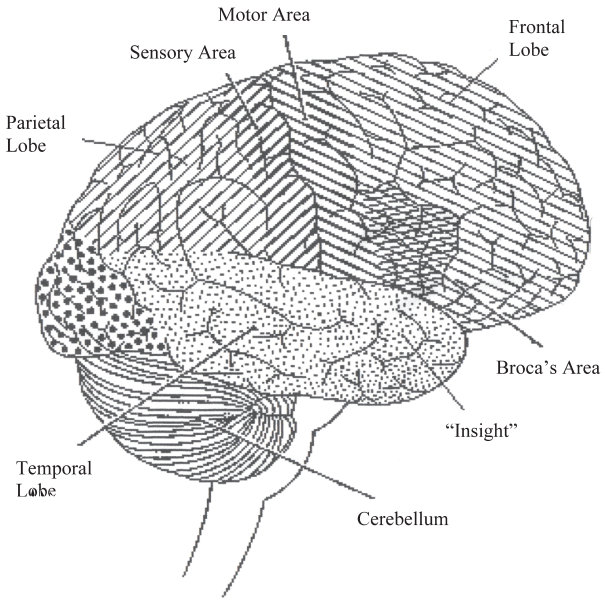

Figure 1. *Lateral view of the human brain (right side). The cerebellum or "hindbrain" contains approximately 100 billion neurons, more than the rest of the entire nervous system (Andersen, Korbo, & Pakkenberg, 1994). The cite of the "Aha" or "insight" experience found by Jung-Beeman et al. (2004) is indicated in the anterior superior portion of the temporal lobe. Note. Figure adapted from National Institutes of Health Publication No. 01–3440a.*

movements and thoughts are identical control objects (control objects are things we intentionally manipulate or imagine manipulating):

> In thought, ideas and concepts are manipulated just as limbs are in movements. There would be no distinction between movement and thought once encoded in the neuronal circuitry of the brain; therefore, both movement and thought can be controlled with the same neural mechanisms. (Ito, 1993, p. 449)

Thus, the control of the components of working memory in solving problems, for example, reading a newspaper, or mentally rearranging the furniture to make room for a new sofa, and so on is not different in neurological principle from the control of limbs in solving problems (e.g., lifting a cup of coffee to the lips, or using the feet, legs, arms, and hands to execute shooting a basketball into a hoop). Both ideas and limbs are control objects. And, as the use of the various thought or limb components is repeated, the cerebellum acts to make the manipulations smoother, faster, more efficient, and more adaptive (e.g., Ito, 1997).

These early foundational arguments suggest that the evolutionary selective advantage of the greatly enlarging cerebellum and the elaboration of its two-way connections with the cerebral cortex was that the cerebellum's computing capacities were being harnessed as an "operating system" (determining the *which*, *when,* and *where* of cerebro-cerebellar information flows) for the evolving, yoked complexities of advanced prehuman and human movement, language, and thought (Leiner & Leiner, 1997). This million or so years of rapid evolution of cerebro-cerebellar circuitry included, of course, operating-system control of the central executive, visuospatial sketchpad and speech loop of human working memory (e.g., Cabeza et al., 2000; Chein et al., 2003). To provide a view of the larger meaning for mental life of this rather rapid evolutionary advance, Leiner and Leiner (1997) have described the computing capacity of neural connections between the human cerebellum and the cerebral cortex (some 40 million nerve tracts—greater than the number of optic nerve tracts) in some detail. In addition, the cerebellum itself contains approximately 100 billion neurons; this is more than the rest of the entire brain (Andersen, Korbo, & Pakkenberg, 1992). The details of this enormous amount of cerebro-cerebellar connectivity are beyond the scope of this article, but it is significant to recognize that the influence of the cerebellum in the linguistic and spatial/temporal control in the internal world of the brain might be considered to rival, in a mental *imagery-enhancing* sense, the influence of the nerve tracts of the visual system on the perceived spatial and temporal dimensions of the external world (see Leiner & Leiner, 1997, pp. 542–547; Leiner et al., 1986, 1989; Schmahmann & Pandya, 1997). It will be helpful to keep these ideas in mind when, later in this article, we examine Einstein's internal world.

Cerebellar Models

Everyday, we are confronted with a myriad of movement and mental situations, some of which may arise unexpectedly. To meet such a variety of situations in a quick, smooth and adaptive manner control models are learned in specialized neural circuitry in the cerebellum (see Footnote 3). The key to the adaptability of these cerebellar models is that they do not learn specific movement and thought patterns but abstract the dynamics of such movement and thought that occurred previously in the same or similar situations (e.g., Ito, 1997). By virtue of this type of abstractive construction, the resulting dynamics models are adaptable to a broad variety of future situations that share the same general state space of the original movement and thoughts (e.g., Ito, 1984a, 1993, 1997). Neurological details describing how cerebellar dynamics models are learned are beyond the scope of this article but can be simplified for our purposes here in the following manner: The abstraction (or "redescription") of activity of the cerebral cortex into cerebellar dynamics models can be described in terms of convergence ratios of incoming movement and thought information (neural connections) onto cerebellar Purkinje cells and their resulting classification of these inputs into distilled patterns of movement and thought—each Purkinje cell is contacted by approximately 200,000 diverse inputs (for a discussion see, e.g., Houk & Wise, 1995). This convergence-driven abstraction of movement/thought information, based upon error signals during repetitive learning, results in abstracted dynamics models (see Ito, 1997). One may think of the resulting state-space of the dynamics model as a three-dimensional problem space of previously learned movements or conceptual thought patterns (like those, e.g., confabulated in repetitive working memory processing; see, e.g., Haruno et al., 1999, 2001; Imamizu et al., 2003; Ito, 1997).

Cerebellar Role in Manipulation of Thought: Conscious and Unconscious Control in Working Memory

In repetitive learning processes, the cerebellum acquires two types of dynamics models, namely, *forward* (predictor) models and *inverse* (controller) dynamics models. A forward model predicts the outcome of a movement or thought; the prediction is confirmed by the sensory consequences of the forward action, either motor or mental. For example, everyone has had the experience of assuming the predictive "hypothesis" that a particular cup of coffee was full when it is actually empty. Then, when the cup is lifted, the discovery is made that the hypothesis was in error as the cup overshoots the expected degree of lift. The point is that the cerebro-cerebellar system had used a forward model of the situation that turned out to be in error. On the other hand, *inverse* dynamics models are defined as neural representations of the motor commands related to control objects necessary to achievement movement goals, just the opposite of the forward models (see Ito, 1997, p. 481; Kawato & Gomi, 1992, p. 445–446). In everyday language, this differentiation means that predictor (forward) models permit behavioral/mental predictions associated with rapid, skilled movement/thought while at a conscious level, whereas cerebellar controller (inverse) models (learned from repetitive forward models) permit the motor cortex to be bypassed, thus allowing rapid, skilled movement/thought to take place at an unconscious level. When we are first learning to shoot baskets, drive a car, or solve problems in working memory (e.g., make change in a foreign currency) each conscious attempt is controlled by a predictor (forward) model, and error signals indicate the appropriateness of the predictor model. After much practice (repetition) and the predictor model is established, a controller (inverse) model is formed as practice continues so that the task can be accomplished automatically without conscious effort. Forward models often generate novel or creative behavior and cognition; inverse models allow the establishment of automaticity. For simplicity, forward and inverse dynamics models will hereafter be referred to as *predictor* and *controller* models.

Ito (1997) provided an example of the respective operation of predictor and controller models that is highly pertinent to the learning of models of working memory activity by the cerebellum:

> According to the psychological concept of a mental model (Johnson-Laird, 1983) [see Footnote 3], thought may be viewed as a process of manipulating a mental model formed in the parietolateral association cortex by commands from the prefrontal association cortex [see Figure 1]. A cerebellar microcomplex [consisting of a Purkinje cell and so forth as mentioned earlier] may be connected to neuronal circuits involved in thought and may represent a dynamics [predictor] or an inverse dynamics [controller] model of a mental model. In other words, a mental model might be transferred from the parietolateral association cortex to the cerebellar microcomplex during repetition of a thought. By analogy to voluntary movement, one may speculate that formation of a dynamics [predictor] model in the cerebellum would enable us to think correctly [and rapidly] in a feedforward manner, i.e., without the need to check the outcome of the thought. This may be the case when one performs a quick arithmetical calculation [accomplished, of course, in working memory]….However, an inverse dynamics [controller] model in the cerebellum would enable us to think automatically without conscious effort. (p. 483)

Ito's speculation that forward mental models in the cerebellum would allow us to think correctly in a feed-forward manner has been born out by more recent research on the cognitive modeling functions of the cerebellum (e.g., Imamizu et al., 2000; Imamizu et al., 2003; Wolpert et al., 2003). It is important to recognize here that Ito is talking about a cerebellar model of a mental model of the cerebral cortex. How many predictor (forward) and controller (unconsciously manipulated) mental models might the cerebellum abstract from working memory processes? And how might these models become related to one another? Such questions are basic to the robustness of the raw material of creativity and innovation.

To begin to answer these questions and to address how they might be related to creativity and innovation, we now examine how predictive and controller cerebellar models are thought to operate in multiple, tightly coupled pairs which can decompose, reorganize, and newly combine aspects of the thought processes of working memory.

Multiple Pairs of Predictor and Controller Cerebellar Models: The Basis of Synthesis in Working Memory That Leads to Creativity and Innovation

Wolpert and Kawato (1998) proposed multiple-paired predictor (forward) and controller (inverse) models for motor and thought-related imagery control.

Within this cerebellar architecture, a relatively small number of separate but interconnected pairs of predictor and controller models have been shown to cover an enormous range of learning and control contexts, including both motor and cognitive tasks such as those associated with working memory (e.g., Haruno et al., 1999; Haruno, et al., 2001; Imamizu et al., 2003; Kawato, 1999). As indicated earlier, this overall adaptive cerebellar architecture is referred to as MOdular Selection and Identification for Control (MOSAIC; e.g., Haruno et al., 2001; Imamizu et al., 2003; Wolpert et al., 2003).

In the MOSAIC cerebellar architecture each of several paired models (a predictor and a controller model) that is brought to bear on a particular behavioral/mental situation has a particular "responsibility" predictor function associated with it. This responsibility predictor determines the pair's contribution to the current movement or thought situation. The responsibility predictor of each pair of models is probabilistic, based upon its history of learning errors in that context. For example, in a case in which working memory is repetitiously culling and mulling over information in a problem situation, the responsibility predictor function would result in the activation of some predictor (forward) models of the several pairs and not others. The activation of these predictor (forward) models would be experienced in working memory as situation-related imagery, the execution of situation-related overt behavior, and would also include imagery associated with stimulus-independent thought (e.g., contemplation, reflection, daydreaming; Teasdale et al., 1995), and "internal speech" (e.g., Ackermann, Mathiak, & Ivry, 2004). As working memory processing continues moment-by-moment, the details of the mental situation would change, and responsibility would accordingly shift to other predictor (forward) models as various pairs' appropriateness to the situation also changed. In addition, when working memory is manipulating different ideas with unknown collective dynamics, the cerebellum simultaneously runs multiple predictor models to test hypotheses concerning their appropriateness to the problem at hand (Wolpert et al., 2003, p. 596). That is, each predictor model can be thought of as a hypothesis tester for the problem situation being addressed in working memory.

The multiple-paired models in the MOSAIC architecture have the following properties that, when combined, are especially salient to the synthesis of both variant and wholly new ideas in working memory. First, the predictor (forward) model of each pair is used to adaptively simulate the imaginary manipulation of thoughts/ideas (Imamizu et al., 2003; Wolpert & Kawato, 1998) in stimulus-independent thought in working memory. In addition, multiple predictor (forward) models can be combined to produce an enormous repertoire of such adaptive mental imagery. The origins of mental imagery in working memory (Baddeley & Andrade, 2000) will be related to such creativity- and innovation-related simulations in the next section. Second, multiple pairs of models can represent decomposed motor/conceptual primitives (Haruno et al., 1999; Imamizu, et al., 2003; Wolpert & Kawato, 1998). This allows complex problems to be broken down into smaller subproblems, each learned by a separate pair of models or collection of paired models. The special relevance of this property of decomposition to creativity and innovation is discussed in terms of Mandler's (2004) work on the origin and nature of conceptual primitives in the next section of this article. Third, each pair of multiple paired models generalizes to novel objects (motor or conceptual primitives) whose dynamics lie within the problem space of already-learned dynamics (Haruno et al., 2001; Wolpert & Kawato, 1998). Thus the responsibility predictor function, along with the above features of the paired-models architecture, acts to explore both previously learned and newly formed problem spaces.

In addition to the above properties of MOSAIC, a person's constellation of models within the MOSAIC architecture is learned over time in hierarchically related layers (Haruno et al., 2003). As learning progresses, concepts are derived on top of established movement and thought repertoires at different levels of abstraction, and a hierarchy of responsibility control for these different levels develops. Wolpert et al. (2003) refer to this hierarchical MOSAIC cerebellar architecture as *HMOSAIC*. Responsibility control for movement and thought can be shifted up and down the learned layers in the HMOSAIC architecture. For example, as will be seen in the next section, image-schematic conceptual primitives learned in early infancy (a foundational layer of HMOSAIC) powerfully influences the development of language modeling in a bottom-up direction (Mandler, 2004), and high-level lan-

guage abstractions learned in adulthood can downwardly influence the ongoing decomposition of these same image-schematic conceptual primitives for use in the visuospatial sketchpad of working memory to solve highly abstract problems (higher level layers of HMOSAIC).

We argue later in the section on Einstein's autobiographical reports of discovery that all of the above properties of paired models within the HMOSAIC architecture allow us, in terms of imagery control, to get inside the constantly shifting imaginary world of creation and innovation in the same way HMOSAIC allows us to get inside the control operations of the worlds of movement and cognition (see Oztop et al., 2005; Wolpert et al., 2003).

On What Theoretical Bases Can Autobiographical Accounts of Creativity and Innovation Be Interpreted?

Before going on to the analyses of Einstein's autobiographical reports, we must address two interrelated theoretical issues that are critically important to understanding the bases upon which we can interpret the creative and innovative mental imagery of working memory and the cerebellum. First, what was the evolutionary role of the progressive collaboration of working memory and the cerebellum in regard to the exploration of novel patterns of behavior and cognition? The working memory/cerebellar mechanisms that evolved to subserve such capacities for novelty, it will be seen, are proposed to be the progenitors, in elaborated form, for creativity and innovation. Second, within this evolutionary context, what was the coadaptive relationship among the three components of working memory, the central executive and its two recursive (rehearsal) slave systems, the speech loop, and the visuospatial sketchpad? Details of the coadaptative evolution of the probable categories of phenomenology (conceptual primitives) of these three components and how they are modulated by the MOSAIC and HMOSAIC cerebellar architectures will allow us to get "inside," so to speak, Einstein's autobiographical accounts of the mental imagery of creativity and discovery.

The Evolutionary Premise for the Collaboration of Working Memory and the Cerebellum in Exploring Novel Situations

The requirements for novel behavior arise in the vagaries of the everyday survival situations of all organisms. Perhaps for highly mobile vertebrates the most common daily selective situations involving the requirement of making fast-paced, novel responses are prey–predator skirmishes. Imagine a wolf chasing a rabbit following zigzag pursuit and escape paths through a complex environment with some moments unexpectedly favorable to either the wolf or the rabbit. These daily skirmishes, of course, are usually either deadly or reaffirming and thus strongly favor the selection of swift, appropriate visuomotor control and memory search and decisional processes. An inherited capacity in the nervous systems of such animals (including humans) to quickly learn and then, without error, execute escape/avoidance responses to novel behavior on the parts of other animals and, at the same time, to also initiate novel behavior of their own in such skirmishes would have been critical to survival.

As part of Leiner et al.'s (1986) foundational arguments (presented earlier in this article) that the cerebellum modulates not only motor behavior but also ideas and language, they suggested how cerebellar modulation of conceptual information would be an advantage in novel situations:

> In confronting a novel situation, the individual may need to carryout some preliminary mental processing before action can be taken, such as processing to estimate the potential consequences of the action before deciding whether to act or to refrain from acting. In such decision-generating processes, the prefrontal cortex is activated (Roland, 1984a). This cortex, via its connections with the cerebellum, could use cerebellar preprogramming to manipulate conceptual data rapidly. As a result, a quick decision could be made. This could then be communicated to the motor areas, including the supplementary motor area (Goldman-Rakic, 1984, p. 448)

Such preliminary mental processing would take place in conjunction with frontal motor imagery areas in which imagined movements are activated (see, e.g., Roland, 1984; Wolpaw, Birbaumer, McFarland, Pfurtscheller, & Vaughan, 2002). In humans, this rapid "preliminary mental processing," simultaneously involving areas of the prefrontal cortex and the cerebellum would clearly involve highly skilled mental devel-

opment among the visuospatial sketchpad, the speech loop and central executive functions of working memory and would be carried out in a combination of conscious and unconscious processes (e.g., Cabeza & Nyberg, 2003; Chein et al., 2003). Of course, the brains of wolves and rabbits mentioned earlier also have the prefrontal cortex, working memory (without the speech loop, obviously), the cerebellum, and so forth.

The above novelty-related principle of highly skilled, rapid "preliminary mental processing" of alternative (creative and innovative) scenarios is precisely what is elaborated, we believe, by the responsibility predictor properties of multiple-paired models in the HMOSAIC cerebellar architecture. In addition, we believe it is what is elaborated on in the long-term working memory development of experts and exceptional performers studied by Ericsson and Kintsch (1995). In support of this contention Ericsson (2002, 2003a, 2003b) has strongly emphasized the finding that, "the essence of expert performance is a generalized skill at successfully meeting the demands of new situations and rapidly adapting to changing conditions" (Ericsson, 2002, p. 41). Ericsson and Kintsch (1995) and Ericsson and Lehmann (1996) found that this generalized skill includes superior planning, reasoning, evaluation of circumstances, and anticipation of future events (see also, Ericsson, 2002, 2003a, 2003b). Moreover, Ericsson and his colleagues have found that the above expert performance "is mediated by complex modifiable representations that allow experts to exhibit faster speed, superior selection of actions, and more precise motor execution" (Ericsson, 2003b, p. 100; and see Footnote 1). All of the behavioral and cognitive capabilities of experts and exceptional performers, learned over many years of deliberate practice (Ericsson, 2002), are precisely what one would expect from the long-term collaboration of working memory and the cerebellum as described by the MOSAIC and HMOSAIC cerebellar architectures described earlier (e.g., Cabeza & Nyberg, 2000; Chein et al., 2003; Desmond & Fiez, 1998; Imamizu et al., 2003; Ito, 1993, 1997; Wolpert et al., 2003).

We hypothesize that it is likely, then, that the evolution of working memory/cerebellar responses to novelty and the manufacture of novelty through the anticipatory mental processing suggested by Leiner et al. (1986) represent the foundational bases of creativity and innovation. Whenever a person confronts a novel problem, whether it be in the context of ancient-era survival situations, a series of novel problems facing Edison as he worked on the telephone or the electric light, or an expert working through the incremental, long-term steps of acquiring exceptional mastery, exploratory cerebellar modulations of cortical activity associated with working memory are the fundamental sources of creative and innovative solutions.

The Coadaptive Evolution of the "Phenomenology" of the Three Components of Working Memory

The dynamic relationships among the three components of working memory were no doubt importantly modified by selective pressures of language evolution that occurred during Pleistocene hominid evolution (cf. Klein et al. 2002). Because the visuospatial sketchpad and the speech loop are proposed to be coevolved slave components of working memory (Baddeley, 1993), it seems that language was likely selected because it allowed more useful sharing of visuospatial imagery that we have in common with prehuman species (including wolves and rabbits). In attempting to understand the relationship between the evolution of the visuospatial sketchpad and the speech loop, then, it seems that the origins and nature of the image structure of the visuospatial sketchpad must be viewed as a precursor to and then underlying structure for the language basis and operation of the speech loop.

The structure of visuospatial/speech imagery and its relationship to language evolution. The important thing about the central executive's use of the activities of visuospatial sketchpad and the speech loop is the *phenomenal imagery* (the imagery of one's mental world) that is produced and kept in a readily accessible state (Baddeley, 1992; Baddeley & Andrade, 2000). This phenomenal imagery constitutes our uniquely human cognitive experience of the world. (How imagery underlies language is described below.) And, because this imagery appears also to be the source of stimulus-independent thought (contemplation, reflection, daydreaming; Christoff & Gabriell, 2000; Singer, 1966; Teasdale et al., 1995), it must be understood if we are to make sense of Einstein's (or anyone else's) contemplative accounts of the discovery process. Baddeley and Andrade (2000) proposed an evolutionary scenario for how such phenomenal imagery may have been selected into working memory:

> Why should working memory play any role in the phenomenology of imagery? Baddeley (1993, 1998, chap. 18) has proposed that working memory plays a central role in the processes underlying consciousness, and that it has evolved as a means of allowing the organism to consider simultaneously a range of sources of information about the world, and uses these processes to set up *mental models* [emphasis added] that facilitate the prediction of events and the planning of action. Consider, for example, the task of a hunter-gatherer who recollects that as this time of year a tree bears fruit near a waterfall in potentially hostile territory. In order to reach the tree safely, he may need to use remembered spatial cues, together with the sound of the waterfall and the shape of the tree, while listening and looking for signs of potential enemies. A dynamic image that is capable of representing these varied sensory features simultaneously is likely to provide a planning aid of considerable evolutionary value. (p. 128)

The important point here is that working memory's control of phenomenal imagery sets up and provides later access to mental models of the meaningfulness of things, for example, "a tree bears fruit," "near a waterfall," "in potentially hostile territory," and so on. It is important to note here that the term *mental model* in Baddeley and Andrade (2000) refers to the same Johnson-Laird mental model referred to earlier in this article by Ito (1997). See Baddeley (1993, 1998, chap. 18) and Footnote 3 of this article.

But how does the central executive of working memory decide what is "meaningful" in terms of survival, and therefore should be built into pertinent models of phenomenal imagery? It cannot reliably obtain this information through the activity of its slave systems because it supervises, integrates, and schedules the activities of the slave systems in the first place (e.g., Baddeley, 1992). So, how does the central executive acquire its *original* rules for constructing phenomenal imagery about what is meaningful?

Working in the area of early conceptual development in infancy, Mandler (1988, 1992a, 1992b, 2004) proposed that perceptual analytic processes occurring during infancy (as early as 3 months) "redescribe" 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 handy synopsis of the tenets of her position:

> The theory proposes that perceptual analysis redescribes perceptual information into *meanings* [emphasis 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 "redescribes" (recodes) perceptual information (both visual and kinesthetic) into spatial meanings and thus initiates the beginnings of concept formation. Mandler (1992b) further proposed that the "redescription" process begins whenever the infant attentively "notices" (not merely looks at) some aspect of the environmental/bodily stimulus array. She indicated that the redescription of perceptual information results in a simplified form of information that is of less detail but of "distilled meaning" (Mandler, 1992b, p. 277). This expression sounds as if the repetitive perceptual analytic process creates cerebellar models, and we believe it does; however, Mandler did not propose brain mechanisms which might account for the redescription process, or how the distilled meanings come about.

The evolutionary structure underlying the phenomenology of working memory. Mandler's (2004, chap. 1) position on the conceptual foundations of an infant's mind included three major points. First, although image schemas themselves are not conscious, they provide the infant with a basis for an accessible conceptual system of imagery that is conscious. Second, image schemas provide structure and meaning to the phenomenal imagery of our thought processes. And, third, image schemas provide a framework of meanings onto which language can be mapped.

Mandler's (1992b, 2004) model can be usefully interpreted as a picture of the evolution and early operation of the *phenomenology* of working memory (as Baddeley & Andrade, 2000, termed it), starting in infancy. In this interpretation, the central executive of working memory arises from the innate biases that first manifest in perceptual analysis. The resulting im-

age-schematic meanings (originally erected through the central executive and cerebellar modeling) provide the neurological platform that drives the imagery of the visuospatial sketchpad. The system of meanings inherent in the visuospatial sketchpad provides a distilled, perceptual/cognitive structure upon which advanced phonological (speech) information and language is mapped (Mandler, 2004). Accordingly, it can be hypothesized from this point of view that the selective evolution of language provided the advantage that the central executive, with cerebellar modulation, could further organize, control, and accelerate the flow of imagery in thought and, of course, in communication. Put a slightly different way, the purpose of language evolution was to control, further decompose and recompose, organize, and speed the flow of imagery in adaptive ways (these are the properties of multiple pairs of cerebellar models within the HMOSAIC architecture listed earlier in this article; e.g., Wolpert et al., 1998, 2003). Language would be represented in a higher level (more abstract) layer of HMOSAIC than the more fundamental image schemas (conceptual primitives) proposed by Mandler. Thus Mandler's image schemas would provide the bottom-up meaning basis for the visuospatial sketchpad in a foundational layer of HMOSAIC, whereas language would act to decompose and reorganize the multiple-paired models of image-schemas in a top-down fashion to construct an unending world of abstract conceptions, while retaining the image-schematic grounding in meanings. Examples of this bidirectional control (Wolpert et al., 2003) in the HMOSAIC architecture is presented in the next section as part of the analysis of Einstein's autobiographic accounts of creativity and discovery.

Einstein's Autobiographical Accounts of Discovery and Invention

Einstein's Inner World

As identified earlier, there are three classic sources that contribute to Einstein's accounts of thinking, imagery, and mathematical discovery: (a) Einstein's responses to Hadamard's (1945) study of invention/discovery in the mathematical field, (b) Einstein's (1949) comments on the nature of thinking that appeared in his "Autobiographical Notes," and (c) his model of the construction of axioms that appeared in a 1952 letter to Maurice Solovine (Einstein, 1956; a thorough theoretical and epistemological discussion of the letter to Solovine appears in Holton, 1979) The above sources extend over several years and provide an excellent sampling of Einstein's persistent views on the workings of his own inner world. All three of the reports contain clear and differentiated meaty points about the structure and dynamics of his discovery-related imagery and are therefore especially amenable to analyses in terms of working memory and the HMOSAIC cerebellar architecture. Along with the discussion of the role of HMOSAIC in each of these analyses, relevant sources that indicate collaborative roles between the cerebellum and pertinent regions of the cerebral cortex in the specific component process of working memory are, to avoid redundancy, indicated only the first time the related component process is mentioned.

A View of the Central Executive

Einstein's earliest account, his responses to Hadamard's (1945), is perhaps our best chance to view the operation of the central executive in Einstein's working memory and how it is modeled in the HMOSAIC cerebellar architecture. The account was in response to a question on the mental methods that lead to invention and discovery in mathematics. The specific survey question read as follows:

> It would be very helpful for the purpose of psychological investigation to know what internal or mental images, what kind of "internal world" mathematicians make use of; whether they are motor, auditory, visual, or mixed, depending on the subject they are studying. (Hadamard, 1945, Appendix I, p. 140)

Einstein answered this question in the following outline manner:

> The words or the language, as they are written or spoken, do not seem to play any role in my mechanism of thought. The psychical entities which seem to serve as elements in thought are certain signs and more or less clear images which can be 'voluntarily' reproduced and combined.
>
> There is, of course, a certain connection between those elements and relevant logical concepts. It is also clear that the desire to arrive finally at logically connected concepts is the emotional basis of this rather vague play with the above mentioned elements. But taken from a psychological viewpoint, this combinatory play seems to be the essential feature in productive thought—before there is any connection with log-

> ical construction in words or other kinds of signs which can be communicated to others.
>
> The above mention elements are, in my case, of visual and some of muscular type. Conventional words or other signs have to be sought for laboriously only in a secondary stage, when the mentioned associative play is sufficiently established and can be reproduced at will.
>
> According to what has been said, the play with the mentioned elements is aimed to be analogous to certain logical connections one is searching for. (Appendix II, pp. 142–143)

Einstein says that, in mathematical discovery, he consciously (voluntarily) plays with combinations of psychical elements but that they are not in the form of language as written or spoken. Here, Einstein is describing how the central executive of his working memory guides the use of "signs" and "images," that can be manipulated in his visuospatial sketchpad. (For attentional cerebellar processes related to the central executive, see Akshoomoff et al., 1997; for cerebellar processes related to spatiotemporal sketchpad, see Lalonde, 1997; Molinari, Petrosini, & Grammaldo, 1997.) The combinatory play (initiated by the central executive) of these "psychical entities" seems to be the key to the overall meaning of Einstein's answers. In HMOSAIC, such combinatory play would be the subjective experience of multiple predictor (forward) models entering consciousness as hypothetical new connections with the "relevant logical concepts" he mentioned. This relevancy would be determined by responsibility functions of predictor (forward) models of previously learned logical concepts operating both within and between layers of MOSAIC within his hierarchical MOSAIC cerebellar architecture (HMOSAIC). It is important to point out here that, while Einstein may not have been aware of "words or language, as they are written or spoken" in his mechanism of thought, they were likely playing a role in the decomposition and reorganization at a higher but silent "internal speech" level in HMOSAIC (see Ackermann et al., 2004).

Einstein responds that conventional words enter only at a secondary stage. This would fit the idea that, in combinatory play, image schemas and logical concepts in lower layers of HMOSAIC are undergoing decomposition and reorganization as Einstein is thinking (and learning). Then, as this process continues, only when a logical concept makes a new, "sufficiently established" connection between layers of HMOSAIC and thus produces a new predictor (forward) and controller (inverse) model pair can it be associated in an upward direction with analogous responsibility signals that would connect it with language control models that reside in a higher level of HMOSAIC (Wolpert et al., 2003).

Finally, Einstein told us how the combinatory play is aimed at being "analogous to certain logical connections one is searching for." What guides this searching process? We believe that the "search" for logical connections is driven by the same processes that drive the construction of Mandler's (2004) spontaneous perceptual analysis in infants. Recall that, according to this view, the conceptual primitives that arise from perceptual analysis are movement-related (e.g., animacy, causality, agency, path) and are thought to represent the basic layer of MOSAIC in the HMOSAIC cerebellar architecture. At the same time, this foundational layer of HMOSAIC contains the earliest models for the control of attention (another way of characterizing "searching") in the central executive (see Akshoomoff et al., 1997, for attentional functions of the cerebellum). This foundational movement-related explanation also helps us understand Einstein's otherwise puzzling inclusion of "muscular type" psychical entities mentioned above. This idea is corroborated by the long-known fact that the cerebellum is importantly involved in motor imagery (e.g., Decety et al., 1994).

In the overall foregoing picture of processing between working memory and HMOSAIC, it can be seen how new paired mental models representing creative and innovative ideas are established in association with decomposed image-schematic conceptual primitives (namely, Einstein's combinatory play of "certain sign and more or less clear images" and not words). In the view presented here, several layers of HMOSAIC guide this process of creative synthesis, but the foundational layer provides its meaning.

According to Einstein, the key to discovery appears to be in the combinatory play with certain "signs and more or less clear images." In the next report from Einstein, we discover more about the structure and dynamics of these "psychical entities."

Visualizing the Visuospatial Sketchpad: Decoding Cerebellar Models

In his 1949 "Autobiographical Notes" Einstein focused on the psychical elements and processing structure of his internal world. He offered a visual record of

the combinatory play and structure that brings forth an *ordering element*. This ordering element provides the logical connections that he mentioned "searching for" in his above response to Hadamard. Einstein referred to the process that unlocks the ordering element of combinatory play as "thinking":

> What, precisely, is "thinking"? When, at the reception of sense-impressions, memory-pictures emerge, this is not yet "thinking." And when such pictures form series, each member of which calls forth another, this too is not yet "thinking." When, however, a certain picture turns up in many such series, then—precisely through such return—it becomes an ordering element for such series, in that it connects series which in themselves are unconnected. Such an element becomes an instrument, a concept. I think that the transition from free association or "dreaming" to thinking is characterized by the more or less dominating role which the "concept" plays in it. It is by no means necessary that a concept must be connected with a sensorily cognizable and reproducible sign (word); but when this is the case thinking becomes by means of that fact communicable. (Einstein, 1949, p. 7)

In Einstein's example of thinking, perhaps in a context of possible equations being used to solve a problem that had previously defied solution, each predictor (forward) model would represent a line of "memory-picture" imagery related to a hypothetical solution. These lines of imagery would include conscious-level decompositions of Mandler's (2004) image schemas as represented in HMOSAIC cerebellar architecture modeling (Wolpert et al., 2003). In Einstein's memory-picture description of thinking these picture series would be activated, by central executive attentional control, in the conscious visuospatial sketchpad of working memory. That is, as in Ito's (1997) earlier quoted example of a predictor (forward) model executing a quick arithmetical calculation, each predictor model would produce a series of memory-pictures in the visuospatial sketchpad of working memory (see also, Wolpert et al.).

Let us say, for example, that there are three sets of such predictor models repetitiously running in Einstein's working memory as he is confronting a problem in his thinking. Each of these three lines of imagery may be thought of as a hypothesis about the solution to the problem situation at hand (see Wolpert et al., p. 596). Then, Einstein says an ordering-element picture emerges from these series. In the HMOSAIC model analysis, this ordering-element picture would be a new "least error" picture arising from the mix of the three predictive (forward) models that is changing in an on-line fashion in Einstein's thinking (Wolpert et al.; see also Imamizu et al., 2003). What would guide the formation of the new least-error predictor (forward) model? Recall, as argued in the last section, that working memory/cerebellar operations are, within the present theory, ultimately based on the system of image-schematic primitives (basic meanings) learned in infancy (Mandler, 2004). Accordingly, the current state of these image-schematic primitives in a lower layer of Einstein's personal HMOSAIC would act to guide the error calculations of his three "competing" hypothetical predictor imageries toward the online construction (by decomposition and generalization) of the new ordering element (see, e.g., Haruno et al., 1999). That is, the "order" in the ordering element would derive its meaning from the image-schematic layer (foundational layer) of HMOSAIC in relation to logical connections already learned in other (higher) layers of HMOSAIC.

The moment of "intuition." Because the predictor (forward) models can be adaptively decomposed and recombined as the mental situation changes, and because the new, ordering-element picture would only be experienced in Einstein's working memory by virtue of a new, separate predictor (forward) model, it would represent the formation of a new pair of predictor/controller models with its own responsibility in that situation (Wolpert et al., 2003; Haruno et al., 1999). In addition, because this newly formed pair of models would be the result of additional rounds of abstractive repetition in working memory, they would operate at higher level of abstraction in HMOSAIC (Wolpert et al.) and thus represent a new, more general concept as described earlier. If this new concept (new ordering element) alters the responsibility predictors of other paired models (ties together other concepts or concept elements) at other levels of Einstein's HMOSAIC, it would represent a new ("creative") idea and be perceived in working memory by him as such. This sudden, unexpected new conceptual knowledge (the new ordering element) connecting different layers of HMOSAIC in a new way would be experienced as an insight. In support of this idea a recent study involving both functional magnetic resonance imaging and electroencephalogram (EEG) methods has shown sudden bursts of neural activity in the anterior temporal lobe at moments of subjective insight involving sudden dis-

covery of new (new to the subject) conceptually connected material (Jung-Beeman et al., 2004).[4] See Figure 1. Following in the vein of these findings, such sudden bursts of neural activity might certainly be expected to occur upon the sudden formation of a new, multilayer concept which might take the form of, for example, insight related to a mathematical axiom (which Einstein, 1956, said are always the result of "intuition"), Kekulé's discovery of the structure of Benzene, or a part of the string of innovative insights leading toward Edison's development of, for example, the telephone or electric light (e.g., Carlson, 2000).[5] Of course, each of these new inspirations of creation and innovation had to be empirically tested, and, as Einstein said in his next account, is always subject to revocation.

The Axiomatic Ordering Element: The Working Memory/Cerebellar Confabulation Leading to Intuition and Insight

Einstein's final report helps clarify how the ordering element, which might turn out to be an axiom in mathematical discovery, starts with experience but is "intuitively" constructed. In a letter to his good friend Maurice Solovine, Einstein (1956) described a complete diagrammatic model of his view of the discovery of axioms. In the model, discovery begins with immediate sensory experience, labeled (E). Then, in the next step, Einstein shows an *intuitive leap* to axioms (A). In the letter he described flow of events in the diagram as follows:

> A are the axioms from which we draw consequences. Psychologically the A are based upon the E. There is, however, no logical path from E to A, but only an intuitive (psychological) connection, which is always subject to "revocation." (p. 121)

Einstein went on to say that once axioms were intuitively conceptualized, logical assertions could be deduced, and that, finally, these assertions must be tested against experience. The idea that axioms could only be arrived at by an intuitive leap was Einstein's most persistent epistemological belief (Holton, 1979).

Within the HMOSAIC modeling architecture, how could someone intuitively know how the mathematical world (or Kekulé's physical world) works before the intuitive idea is actually tested? We propose that the same working memory/HMOSAIC arguments that applied to the construction of the ordering element (new concept) in Einstein's definition of "thinking" apply to the workability of intuition in the real physical world.

Within this view, intuitions are sometimes workable in the physical world, because the neurological principles behind the operation of the image-schematic base of working memory and HMOSAIC predictor (forward) models act as biases toward the principles that govern the physical world. This predictor model bias toward physical world principles originates in the perceptual analysis of environmental dynamics in infancy Mandler (2004) and its probable anchoring in the mapping functions of the brain's hippocampus. (For a discussion of the connections between the cerebellum and visuospatial functions of the hippocampus see, e.g, Lalonde, 1997; Molinari et al., 1997; Vandervert, 1997.)

Conclusion

We have attempted to bring together the substantial bodies of research on the neurophysiology of working memory and the cognitive functions of the

[4]Although Jung-Beeman et al.'s (2004) results are helpful to understanding the location in the brain (temporal lobe) of the common "aha," or "eureka" experience associated with creative discovery, they do not shed light on the processes that lead to concept formation and insight, for example, those processes mentioned by Einstein in his three subjective accounts. The "aha" report measured by Jung-Beeman et al. can only be interpreted as a brief signal that all of the "creative" work (completion of a Remote Associates Test [Mednick, 1962] type item) had been completed.

[5]The proposed occurrence of insight upon the merging of layers of the HMOSAIC cerebellar architecture is reminiscent of Koestler's (1964) concept of creative "bisociation:"

> I have coined the term 'bisociation' in order to make a distinction between the routine skills of thinking on a single "plane," as it were, and the creative act, which, as I shall try to show, always operates on more than one plane. (pp. 35–36)

Elsewhere, he elaborated on the idea:

> The Latin *cogito* comes from *coagitare*, to shake together. *Bisociation means combining two hitherto unrelated cognitive matrices in such a way that a new level is added to the hierarchy, which contains the previously separate structures as its members*. (Koestler, 1967, p. 183)

cerebellum to propose a new theory of innovation and creativity. In addition, we have applied this theory to an analysis of autobiographical accounts of creativity and innovation. This type of analysis provides a detailed, phenomenally experienced counterpart to behavioral, clinical, and neuroimaging studies that further elucidates correlations between working memory and cognitive functions of the cerebellum. The value of an analysis of autobiographical accounts of working memory is that the mapping of certain categories of the phenomenal imagery of creativity and innovation to particular components of working memory and modularity within the cerebellum can provide a theoretical model that can suggest where the activity of creativity and innovation might occur in the brain. Such information will help design future experiments aimed at locating creativity and innovation, in vivo, in specific complexes of cerebro-cerebellar circuitry. In order to achieve a better understanding of this circuitry, we must determine not only the locations of the neural sources but also their temporal relationships. Although functional neuroimaging techniques can identify active brain regions, they are not able to disclose the temporal courses of the circuitry. This information can be obtained through analysis of the electromagnetic signals of the brain such as EEG, which has millisecond temporal resolution. Recent progress in obtaining a non-invasive, integrated spatial/temporal analysis of neural activity has come in the form of advances in inverse methods, which bring the power of spatial cortical mapping to the electromagnetic techniques (Pascual-Marqui, Michel, & Lehmann, 1994; Mosher & Leahy, 1998; Liu & Schimpf, in press; Liu et al., in press). By combining functional neuroimaging and electromagnetic techniques (Dale et al., 2000), the circuitry devoted to creative and innovative activity may be measured as it unfolds. In an effort to further validate the respective roles of working memory and the cerebellum in creativity and innovation we are currently investigating a variety of brain areas during sustained in vivo creative and innovative tasks.

References

Ackermann, H., Mathiak, K., & Ivry, R.B. (2004). Temporal organization of "internal speech" as a basis for cerebellar modulation of cognitive functions. *Behavioral and Cognitive Neuroscience Reviews, 3,* 14–22.

Akshoomoff, N., Courchesne, E., & Townsend, J. (1997). Attention coordination and anticipatory control. In J.D. Schmahmann (Ed.), *The cerebellum and cognition* (pp. 575–598). New York: Academic Press.

Andersen, B., Korbo, L., & Pakkenberg, B. (1992). A quantitative study of the human cerebellum with unbiased stereological techniques. *The Journal of Comparative Neurology, 326,* 549–560.

Baddeley, A. (1992, January 31). Working memory. *Science, 255,* 556–559.

Baddeley, A. (1993). Working memory and conscious awareness. In A. Collins, S. Gathercole, M. Conway, & P. Morris (Eds.), *Theories of memory* (pp. 11–28). Mahwah, NJ: Lawrence Erlbaum Associates.

Baddeley, A. (1998). *Human memory: Theory and practice*. Needham Heights, MA: Allyn & Bacon.

Baddeley, A., & Andrade, J. (1998). Working memory and consciousness: An empirical approach. In M. Conway, S. Gathercole, & C. Cornoldi (Eds.), *Theories of memory: Volume 2* (pp. 1–23). Hove, England: Psychology Press.

Baddeley, A., & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General, 129*, 126–145.

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

Baddeley, A., & Logie, R.H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). New York: Cambridge University Press.

Bloedel, J.R., Dichgans, J., & Precht, W. (1985). *Cerebellar functions*. Berlin: Springer-Verlag.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience, 12(1),* 1–47.

Carlson, W. B. (2000). Invention and evolution: The case of Edison's sketches of the telephone. In J. Ziman (Ed.), *Technological innovation as an evolutionary process* (pp. 137–158). Cambridge, England: Cambridge University Press.

Chein, J.M., Ravizza, S.M., & Fiez, J.A. (2003). Using neuroimaging to evaluate models of working memory and their implications for language processing. *Journal of Neurolinguistics, 16,* 315–339.

Christoff, K., & Gabriell, J.D.E. (2000). The frontopolar cortex and human cognition: Evidence for a rostral hierarchical organization within the human prefrontal cortex. *Psychobiology, 28,* 168–186.

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

Craik, K. (1943). *The nature of explanation*. Cambridge: Cambridge University Press.

Dale, A. M., Liu, A. K. Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren, E. (2000). Dynamic statistical para-

metric neurotechnique mapping: Combining fMRI and MEG for high-resolution image of cortical activity. *Neuron, 26,* 55–67.

Desmond, J., & Fiez, J. (1998). Neuroimaging studies of the cerebellum: Language, learning and memory. *Trends in Cognitive Sciences, 2,* 355–362.

Decety, J., Perani, D., Jeannerod, M., Bettinardi, V., Tadary, B., Woods, R., et al. (1994). Mapping motor representations with positron emission tomography. *Nature, 371,* 600–602.

Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia, and the cerebral cortex? *Neural Networks, 12,* 961–974.

Einstein, A. (1949). Autobiographical notes. In A. Schillp (Ed.), *Albert Einstein: Philosopher-scientist* (Vol. 1, pp. 1–95). La Salle, IL: Open Court.

Einstein, A. (1954). *Ideas and opinions*. New York: Crown Publishers.

Einstein, A. (1956). *Lettres à Maurice Solovine*. Paris: Gauthier-Villars.

Ericsson, K. A. (2002). Attaining excellence through deliberate practice: Insights from the study of expert performance. In M. Ferrari (Ed.), *The pursuit of excellence through education* (pp. 21–55). Mahwah, NJ: Lawrence Erlbaum Associates.

Ericsson, K. A. (2003a). The acquisition of expert performance as problem solving. In J. E. Davidson & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 31–83). Cambridge, England: Cambridge University Press.

Ericsson, K. A. (2003b). The search for general abilities and basic capacities: Theoretical implications from the modifiability and complexity of mechanisms mediating expert performance. In R. J. Sternberg & E. I. Grigorenko (Eds.), *The psychology of abilities, competencies, and expertise* (pp. 93–125). Cambridge, England: Cambridge University Press.

Ericsson, K. A., & Delaney, P. F. (1999). Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 257–297). New York: Cambridge University Press.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102,* 211–245.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence on maximal adaptations on task constraints. *Annual Review of Psychology, 47,* 273–305.

Fiez, J., Raife, E., Balota, D., Schwarz, J., Raichle, M., & Petersen, S. (1996). A positron emission tomography study of the short-term maintenance of verbal information. *The Journal of Neuroscience, 16,* 808–822.

Goldman-Rakic, P. S. (1992, September). Working memory and the mind. *Scientific American, 267,* 111–117.

Hadamard, J. (1945). *The psychology of invention in the mathematical field*. New York: Dover.

Haruno, M., Wolpert, D., & Kawato, M. (1999). Multiple paired forward-inverse models for human motor learning and control. In M. S. Kearns, S. A. Solla, & D. A. Cohn (Eds.), *Advances in neural information processing systems* (pp. 31–37). Cambridge, MA: MIT Press.

Haruno, M., Wolpert, D., & Kawato, M. (2001). MOSAIC model for sensorimotor and learning control. *Neural Computation, 13,* 2201–2220.

Holton, G. (1979). Constructing a theory: Einstein's model. *The American Scholar, 48,* 309–339.

Houk, J., & Wise, S. (1995). Distributed modular architectures linking basal ganglia, cerebellum, and cerebral cortex: Their role in planning and controlling action. *Cerebral Cortex, 2,* 95–110.

Imamizu, H., Kuroda, T., Miyauchi, S., Yoshioka, T., & Kawato, M. (2003). Modular organization of internal models of tools in the cerebellum. *Proceedings of the National Academy of Science, 100,* 5461–5466.

Imamizu, H., Miyauchi, S., Tamada, T., Sasaki, Y., Takino, R., Pütz, B., et al. (2000). Human cerebellar activity reflecting an acquired internal model of a new tool. *Nature, 403,* 192–195.

Ito, M. (1984a). *The cerebellum and neural control*. New York: Raven Press.

Ito, M. (1984b). Is the cerebellum really a computer? *Trends in Neurosciences, 2,* 122–126.

Ito, M. (1993). Movement and thought: Identical control mechanisms by the cerebellum. *Trends in Neurosciences, 16,* 448–450.

Ito, M. (1997). Cerebellar microcomplexes. In J. D. Schmahmann (Ed.), *The cerebellum and cognition* (pp. 475–487). New York: Academic Press.

Ivry, R. (1997). Cerebellar timing systems. In J. D. Schmahmann (Ed.), *The cerebellum and cognition* (pp. 555–573). New York: Academic Press.

Johnson-Laird, P. (1983). *Mental models*. New York: Cambridge University Press.

Jung-Beeman, M., Bowden, E., Haberman, J., Frymiare, J., Arambel-Liu, S., Greenblatt, R., et al. (2004). Neural activity when people solve verbal problems with insight. *PLOS Biology, 2,* 500–510.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology, 9,* 718–727.

Kawato, M., & Gomi, H. (1992). The cerebellum and VOR/OKR learning models. *Trends in Neuroscience, 15,* 445–453.

Kihlstrom, J. (1987). The cognitive unconscious. *Science, 237,* 1445–1452.

Klein, S. B., Cosmides, L., Tooby, J., & Chance, S. (2002). Decisions and the evolution of memory: Multiple systems, multiple functions. *Psychological Review, 109,* 306–329.

Koestler, A. (1964). *The act of creation*. New York: MacMillan.

Koestler, A. (1967). *The ghost in the machine*. New York: MacMillan.

Kornhuber, H. (1974). Cerebral cortex, cerebellum, and basal ganglia: An introduction to their motor functions. In F. Schmitt & F. Worden (Eds.), *The neurosciences: Third study program* (pp. 267–280). Cambridge, MA: MIT Press.

Lalonde, R. (1997). Visuospatial abilities. In J. D. Schmahmann (Ed.), *The cerebellum and cognition* (pp. 191–215). New York: Academic Press.

Leiner, H., & Leiner, A. (1997). How fibers subserve computing capabilities: Similarities between brains and machines. In J. D. Schmahmann (Ed.), *The cerebellum and cognition* (pp. 535–553). New York: Academic Press.

Leiner, H., Leiner, A., & Dow, R. (1986). Does the cerebellum contribute to mental skills? *Behavioral Neuroscience, 100,* 443–454.

Leiner, H., Leiner, A., & Dow, R. (1989). Reappraising the cerebellum: What does the hindbrain contribute to the forebrain? *Behavioral Neuroscience, 103,* 998–1008.

Liu, H., & Schimpf, P. (2006). Efficient localization of synchronous EEG source activities using a modified RAP-MUSIC algorithm. *IEEE Transactions in Biomedical Engineering, 53,* 652–661.

Liu, H., Schimpf, P., Dong, G., Gao, X., Yang, F., & Gao, S. (2005). Standardized shrinking LORETA-FOCUSS (SSLOFO): A new algorithm for spatio-temporal EEG source reconstruction. *IEEE Transactions in Biomedical Engineering, 52,* 1681–1691.

Mandler, J. (1988). How to build a baby: On the development of an accessible representational system. *Cognitive Development, 3,* 113–136.

Mandler, J. (1992a). The foundations of conceptual thought in infancy. *Cognitive Development, 7,* 273–285.

Mandler, J. (1992b). How to build a baby II: Conceptual primitives. *Psychological Review, 99,* 587–604.

Mandler, J. (2004). *The foundations of mind: Origins of conceptual thought*. Oxford, England: Oxford University Press.

Mednick, S.A. (1962). The associative basis of the creative process. *Psychological Review, 69,* 220–232.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambridge University Press.

Molinari, M., Petrosini, L., & Grammaldo, L. (1997). Spatial event processing. In J. D. Schmahmann (Ed.), *The cerebellum and cognition* (pp. 317–230). New York: Academic Press.

Mosher, J. C., & Leahy, R. M. (1998). Recursive MUSIC: A Framework for EEG and MEG source localization. *IEEE Transactions in Biomedical Engineering, 45,* 1342–1354.

Oztop, E., Wolpert, D., & Kawato, M. (2005). Mental state inference using visual control parameters. *Cognitive Brain Research, 22,* 129–151.

Pascual-Marqui, R. D., Michel, C. M., & Lehmann, D. (1994). Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *International Journal of Psychophysiology, 18,* 49–65.

Roland, P. E. (1984). Organization of motor control by the normal human brain. *Human Neurobiology, 2,* 205–216.

Schmahmann, J. (Ed.). (1997). *The cerebellum and cognition*. New York: Academic Press.

Schmahmann, J., & Pandya, D. (1997). The cerebrocerebellar system. In J. D. Schmahmann (Ed.), *The cerebellum and cognition* (pp. 31–60). New York: Academic Press.

Singer, J. L. (1966). *Daydreaming*. New York: Random House.

Teasdale, J., Dritschel, B., Taylor, M., Proctor, L., Lloyd, C., Nimmo-Smith, I., & Baddeley, A. (1995). Stimulus-independent thought depends on central executive resources. *Memory & Cognition, 23,* 551–559.

Thach, W. T. (1996). On the specific role of the cerebellum in motor learning and cognition: Clues from PET activation and lesion studies in man. *Behavioral and Brain Sciences, 19(3),* 411–431.

Vandervert, L. (1997). The evolution of Mandler's conceptual primitives (image schemas) as neural mechanisms for space-time simulation structures. *New Ideas in Psychology, 15,* 105–123.

Vandervert, L. (2003a). How working memory and cognitive modeling functions of the cerebellum contribute to discoveries in mathematics. *New Ideas in Psychology, 21,* 159–175.

Vandervert, L. (2003b). The neurophysiological basis of innovation. In L. V. Shavinina (Ed.), *The international handbook on innovation* (pp. 17–30). Oxford, England: Elsevier Science.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology, 113,* 767–791.

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London B, 358,* 593–602.